

AD-A105 753

PITTSBURGH UNIV PA LEARNING RESEARCH AND DEVELOPMEN--ETC F/8 6/5
KNOWLEDGE BASED COMPONENTS OF EXPERTISE IN MEDICAL DIAGNOSIS.(U)
SEP 81 P J FELTOVICH
LRDC-81/PDS-2

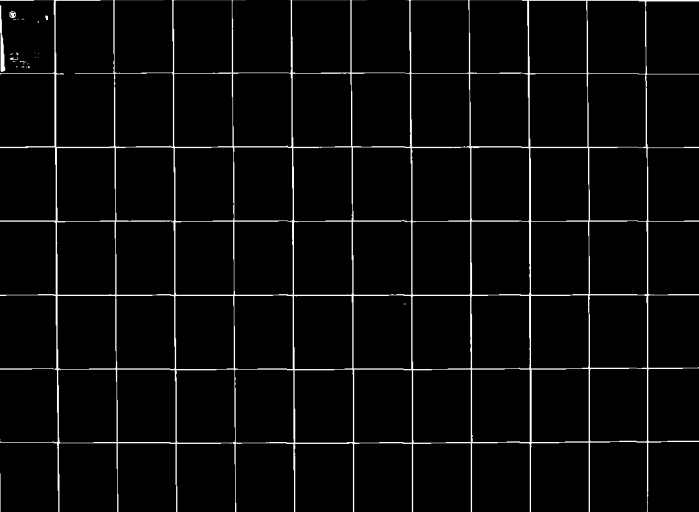
N00014-79-C-0215

NL

UNCLASSIFIED

1 of 3

AD-A105 753



LEVEL II

(12)

AD A105753



**University of Pittsburgh
LEARNING RESEARCH
AND DEVELOPMENT CENTER**

**DTIC
ELECTE
OCT 14 1981
S D H**

**KNOWLEDGE BASED COMPONENTS OF
EXPERTISE IN MEDICAL DIAGNOSIS**

**Paul J. Feltovich
Learning Research and Development Center
University of Pittsburgh**

September 1981

Technical Report No. PDS-2

This research was sponsored by the Personnel and Training Research Program, Psychological Sciences Division, Office of Naval Research, under Contract No. N00014-79-C-0215. Contract Authority Identification Number, NR 797-430.

This report is issued by the Learning Research and Development Center, supported in part by funds from the National Institute of Education (NIE), U.S. Department of Education.

Reproduction of this report is authorized for any purpose of the U.S. Government.

Reference to trade names does not imply endorsement.

U.S. GOVERNMENT

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Technical Report No. PDS-2	2. GOVT ACCESSION NO. AD-A165 753	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Knowledge Based Components of Expertise in Medical Diagnosis.	5. TYPE OF REPORT & PERIOD COVERED Technical Report. 71 Jan 79-30 Aug 81	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Paul J. Feltovich	8. CONTRACT OR GRANT NUMBER(s) N0014-79-C-0215	
9. PERFORMING ORGANIZATION NAME AND ADDRESS Learning Research and Development Center University of Pittsburgh Pittsburgh, Pennsylvania 15260	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 157-430	
11. CONTROLLING OFFICE NAME AND ADDRESS Personnel and Training Research Programs Office of Naval Research (Code 458) Arlington, VA 22217	12. REPORT DATE September 1981	13. NUMBER OF PAGES 265
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report) Unclassified	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Expertise, skill, human performance, skill acquisition, diagnosis, troubleshooting, medicine.		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) The report investigates the contribution of case-related medical knowledge to clinical diagnosis and differences in this knowledge among individuals with different amounts of experience in a subspecialty of medicine. Subjects diagnosed clinical cases while thinking aloud. Each case was designed to assess a different aspect of medical knowledge. Consistent differences in performance among diagnosticians at different levels of experience were found and inferences made to sources of medical knowledge responsible for		

DD FORM 1 JAN 73 1473

EDITION OF 1 NOV 68 IS OBSOLETE
S/N 0102-LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. performance. Recurrent sources of error (bugs) were identified for the less experienced diagnosticians.

ACKNOWLEDGMENTS

This report is based on a doctoral dissertation submitted to the Graduate School of the University of Minnesota. Fundamental contributions were made by my advisor for the work, Paul E. Johnson of the Center for Research in Human Learning, and by James H. Moller of the University of Minnesota Medical School. The contributions by other committee members, Norman L. Chervany, William P. Fox, Thomas J. Hummel, and Herbert L. Pick and by David B. Swanson of the American Board of Internal Medicine are also appreciated. Gratitude is also extended to the physicians who participated in the study and to the Division of Pediatric Cardiology of the University of Minnesota Medical School. The study was completed during my residence at the Learning Research and Development Center, University of Pittsburgh, where my appointment was supported in part by the Office of Naval Research. The paper has benefitted greatly from my collaboration with Alan M. Lesgold, Robert Glaser, and Michelene T. H. Chi. Portions of the study were reported at the annual meeting of the American Educational Research Association, Boston, 1980.

Accession For	<input checked="checked" type="checkbox"/>
NTIS GRA&I	<input type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	<input type="checkbox"/>
For	<input type="checkbox"/>
Distribution/	<input type="checkbox"/>
Availability Codes	<input type="checkbox"/>
Avail and/or	<input type="checkbox"/>
Dist	<input type="checkbox"/>
A	<input type="checkbox"/>

ABSTRACT

The report investigates the contribution of case-related medical knowledge to clinical diagnosis and differences in this knowledge among individuals with different amounts of experience in a subspecialty of medicine. Subjects diagnosed clinical cases while thinking aloud. Each case was designed to assess a different aspect of medical knowledge. Consistent differences in performance among diagnosticians at different levels of experience were found and inferences made to sources of medical knowledge responsible for performance. Recurrent sources of error (bugs) were identified for the less experienced diagnosticians.

TABLE OF CONTENTS

CHAPTER ONE: Introduction	1
CHAPTER TWO: Method	22
Materials	22
Subjects	37
Procedure	41
Data and Analysis	45
CHAPTER THREE: Results	50
Case 1 (Subvalvular Aortic Stenosis)	56
Case 2 (Total Anomalous Pulmonary Venous Connection)	76
Case 3 (Patent Ductus Arteriosus)	91
Case 4 (Pulmonary Atresia)	96
Case 5 (Coarctation of the Aorta plus)	111
Summary of Results	137
CHAPTER FOUR: Discussion	152
BIBLIOGRAPHY	172
Appendix A	183
Appendix B	184
Appendix C	202
Appendix D	206
Appendix E	213
Appendix F	216
Appendix G	231
Appendix H	233
Appendix I	249
Appendix J	261

Knowledge Based Components of Expertise in Medical Diagnosis

1. INTRODUCTION

The present research investigates the effects of medical knowledge on clinical diagnosis and the differences in such knowledge possessed by individuals with different experience in the task of medical diagnosis and with a subspecialty of medicine. The study is useful because questions exist among medical practitioners concerning the relative contribution to diagnostic ability of medical knowledge in comparison to skills of clinical reasoning or "problem solving skills" (e.g., Barrows, 1979). In this regard, the study complements a recent report which demonstrated the importance of medical knowledge in diagnosis but addressed this issue less directly (Lesgold, Feltovich, Glaser, & Wang, 1981). The present work also contributes more generally to the theory of problem solving where the role of domain knowledge has recently gained emphasis (Greeno, 1980).

The Importance of Knowledge in Cognitive Skills

In recent studies distinctions between knowledge and cognitive skills have blurred. That knowledge affects the quality and nature of reasoning, problem solving, and other cognitive skills has been demonstrated. That these skills use knowledge as a substrate appears evident, and even the idea that reasoning is embedded within forms of knowledge has

been advanced.

Recent laboratory research has indicated that knowledge contributes to even the most fundamental cognitive skills. Glaser and others, investigating basic skills of human intelligence such as induction, have suggested that even these are strongly dependent on a person's conceptual knowledge of the domain (e.g., conceptual knowledge of numbers in number analogy and number series tasks) to which the intellectual skill is applied (Corsale & Gitomer, 1979; Glaser & Pellegrino, in press; Pellegrino, Chi, & Majetic, 1978). An indication of the importance of knowledge in skills traditionally considered to measure intelligence is that of the three major components of analogical reasoning proposed by Glaser & Pellegrino (in press) two are directly knowledge related.

The knowledge base possessed by an individual has also been shown to influence fundamental mechanisms of learning, for example, the spontaneous use by subjects of memory strategies such as grouping and rehearsal (e.g., Chi, 1978; Ornstein & Corsale, in press), the ability to use such strategies even under experimental prompting (Chi, 1979), and the amount of information that can be held in short-term memory (Chi, 1978). Voss and colleagues (Chiesi, Spilich & Voss, 1979;

Spilich, Vesonder, Chiesi, & Voss, 1979) have extended work of this sort beyond basic memory tasks into domains of subject-matter learning. Within a given subject matter, high-knowledge individuals have greater recognition and recall memory for new material than do low-knowledge individuals, can make useful inferences from smaller amounts of partial information, and are better able to integrate new material within a coherent and interconnected framework of knowledge (organized, for example, around a common goal structure).

Reasoning, itself, has been shown to be highly dependent on an individual's knowledge base for the task environment in which the reasoning occurs. Wason and Johnson-Laird (1972) present considerable evidence that individuals perform poorly in testing the implications of logical inference rules (e.g., if p then q) when these rules are stated abstractly. However, performance is greatly improved for concrete instances of the same rules (e.g., every time I go to Manchester, I travel by train). Rumelhart (1979) reports an extension of this work in which nearly five times as many subjects were able to test appropriately the implications of a simple, single-conditional logical expression

when the expression was couched in terms of a familiar setting (e.g., a work setting; every purchase for over thirty dollars must be approved by the regional manager) versus when the expression was stated in an understandable but less richly semantic form (e.g., Every card with a vowel on the front must have an integer on the back).

An explanation for this context sensitivity of reasoning ability is that particular situations engage in an individual an infrastructure of related knowledge concerning such things as characteristics of entities or people involved, models of causality or temporal sequence, conventions of conduct, or even records of personal involvement and activity in situations of a similar type. These extralogical knowledge factors dominate formal rules in reasoning. When, for a given setting, these other factors are consistent with formal logic, a person will appear formally rational (In the Rumelhart, 1979, "regional manager" task mentioned above, a rationale consisting of: "So what if the manager signs a few more forms than he really needs to" is enough to exclude from this task the most prevalent type of error subjects make on the companion vowel-integer task, that is, testing that cards with an integer on the back have a vowel on the front).

In situations where extralogical knowledge is inconsistent with formal reasoning or, perhaps more convincingly, where its contribution is taken away altogether as in the vowel-integer task, the relatively anemic nature of content-free, "pure" reasoning is exposed.

Under the assumption that, in the long run, nature rewards logically accurate and punishes logically inaccurate reasoning, one might expect that as individuals acquire extensive experience functioning in particular task environments, their related knowledge will be shaped along logically appropriate lines. Convergent evidence that the resultant reasoning proficiency is highly situation-specific again comes from Wason (Wason & Johnson-Laird, 1972) who reports very little transfer from inference training in one context to proficiency in others. This content-constrained conception of formal reasoning is in contrast to structural developmental theories (e.g., Piaget, 1972) which claim cross-situational, content free, and maturational-ly determined, general reasoning skills. Yet, even within these theories, evidence is emerging for the import of accumulated knowledge as a contributor to these abilities (e.g., Carey, 1973).

These developments in the psychology of reasoning have been mirrored in artificial intelligence research which has shown an evolution from systems in which knowledge (declarative) and reasoning (procedures) were clearly separated to systems in which these components strongly interact or are indistinct. Early systems such as Green's QA3 (Green, 1969) and Quillian's TLC (Quillian, 1969) relied on data bases of uniformly formatted declarative knowledge and a few general purpose reasoning algorithms for operating on these knowledge bases. These systems have given way to ones in which the separation between knowledge and reasoning components is less distinct and in which general reasoning algorithms have considerably less status in comparison to narrowly applicable reasoning strategies embedded in procedures for operating within specific domains of knowledge (e.g., Norman, Rumelhart & LNR, 1975; Sacerdoti, 1977; Van Lehn & Seely Brown, 1979). Again, reasoning is treated not so much as a general but as a task and content specific skill.

Research in problem solving has shown a similar evolution from an emphasis on generality to relative task-specificity and its relationship to bodies of knowledge. Early problem solving theories proposed quite general domain-independent problem solving methods for example,

"means-ends analysis," that were envisioned sufficient to capture cross-situational problem solving abilities (e.g., Ernest & Newell, 1969; Fikes, 1969). Such theories fared reasonably well so long as problem solving was restricted to domains (e.g., puzzles) which involved little domain knowledge and in which the primary obstacle to successful problem solving involved determining the appropriate sequence for a small number of state-changing operators (e.g., Newell & Simon, 1972). As problem solving researchers have addressed tasks more like those encountered in the professions and for which proficiency requires years of training and learning, the critical role of domain knowledge has been recognized. Recent major problem solving systems depend on extensive stores of knowledge both about the particular problem solving domain and particular problem solving strategies effective in the domain (e.g., Buchanan & Feigenbaum, 1978; Friedland, 1979; Shortliffe, 1976; Stevens, Collins, & Goldin, 1979). Such systems have come to be known as "knowledge-based systems" (Barnett, 1977) and the enterprise of harnessing large-scale knowledge bases within these systems for the purpose of solving complex problems has been termed "knowledge engineering" (Feigenbaum, 1977).

Knowledge and Expert-Novice Differences in Problem Solving

The role of knowledge and its organization have been implicated in recent work on expertise and expert-novice differences in problem solving within complex domains. Perhaps the best established characteristic of expert problem solvers is their ability to recognize quickly meaningful events in their problem solving environment. The pioneering demonstrations of this phenomenon were for chess experts and were conducted by deGroot (1965) and later by Simon and Chase (Chase & Simon, 1973a; Simon & Chase, 1973). These investigators found that, after very brief exposure to chess boards extracted from real games, chess experts were able to reproduce much more of a board than were novices. This ability was not due to general superiority of memory, since experts reproduced random boards no better than novices, but rather to the experts' ability to see entire configurations of pieces as single units or "chunks". Expert perceptual and cognitive chunking has been replicated many times in chess (e.g., Frey & Adelman, 1976; Goldin, 1978), in a wide variety of other problem solving fields (e.g., Charness, 1979, in bridge; Egan & Schwartz, 1979, in electronics; Reitman, 1976, the game of GO), and is not limited to visually loaded

tasks (Chase & Chi, 1980)..

The usual explanation for the expert's recognition ability is that with experience experts establish in long-term memory a very large "vocabulary" of memory structures, each representing a recurring problem solving pattern or event, which can then be used to encode subsequent problem solving situations (e.g., Chase & Simon, 1973a,b). Novices, with less exposure to recurrent patterns, have a much smaller store of familiar memory structures, and are more often forced to deal with problems in a novel and piecemeal manner. Simon and Chase speculated that expert memory was organized hierarchically (i.e., specific representations embedded within layers of more general representations) but had no direct evidence for this claim. More recent research has contributed more direct evidence for the hierarchical nature of expert memory utilized in the recognition of problem situations (Akin, 1980; Egan & Schwartz, 1979).

How does expert recognition memory contribute to problem solving? A plausible interpretation of prior research results is that for experts, memory representations for familiar problem settings have associated with them corresponding sequences of good actions. One

source of evidence for this interpretation comes again from chess research which has shown that experts consider no more alternative moves from a given position than do novices, nor do they investigate more consequences of any particular move; experts simply consider better moves. The expert recognizes the situation and calls forth actions that have proved efficacious in the same or analogous situations. Other evidence for memory representations driving solution, comes from recent research in physics problem solving. Larkin, 1980, has proposed a construct of "chunked procedures" to account for the fact that expert solvers generate solution equations in grouped "bursts" while more novice solvers generate equations in a more sporadic and isolated manner. A chunked procedure is a relatively integrated solution plan associated with expert categorization or typing of a problem. Similarly, Chi, Feltovich, & Glaser, 1980, have shown that differences in problem solving processes of expert and novice physics problem solvers result both from differences in the internal structure of memory representations for problem types and from differences in memory organization among these types. While their work has not addressed expert-novice differences, Simon and colleagues (Hinsley, Hayes & Simon, 1976; Paige & Simon, 1966) have shown that schemata,

knowledge structures representing problem types, also strongly influence the nature of problem solving processes in algebra.

A characterization of expertise in problem solving that follows from the findings outlined above is that expertise is largely a matter of the content and organization of knowledge in long term-memory and mechanisms for engaging appropriate knowledge in situations where it is needed. This characterization of expertise, that is, a large, organized long-term memory vocabulary of problem representations, mechanisms for mapping problem events into these representations, and associated courses of action, was at one time proposed for expertise by Newell and termed the "big switch" (Newell, 1973). It has since gained empirical support.

A final characteristic of expertise and expert problem solving that needs to be addressed is its task-specificity. There appears to be little transfer from high-level proficiency in one domain to proficiency in other domains - even when the domains seem, intuitively, very similar. For example, in tasks similar to those used in the chess board studies, Eisenstadt and Kareev, 1975, have studied the memory for brief displays of expert GO and Gomoku players. Even though these two

games are played on the same board and with the same pieces, GO players showed quite poor performance on Gomoku displays and vice versa. In another expert-novice study, some of the counter intuitive results of Thorndyke and Stasz, 1980, are explained by the fact that some of their designated experts were slightly out of their realm, for example, an individual proficient in working with high altitude aerial maps involved in a task of learning city street maps. Task specificity is also characteristic of motor skills (e.g., Pitts & Posner, 1967; Martenink, 1974), and to the extent that motor and cognitive skill development are similar (for example, they both share the same learning curve - Newell and Rosenbloom, in press), this can be taken as further evidence for the task specificity of cognitive skills such as problem solving. Task specificity is what one would expect if, in fact, high-level skill development is largely a matter of knowledge base development (c.f. Chase & Chi, 1980).

The literature on cognitive skills, reasoning, problem solving and expertise overviewed thus far has several implications for the present research on diagnostic expertise. First, the medical knowledge base is likely to constitute a major component of expertise,

and differences in this knowledge base between highly experienced diagnosticians and those less experienced should largely account for differences in proficiency. The study must have means for assessing knowledge components in detail - for example, the ways diagnosticians partition their problem space into categories or types of subproblems and the effects of these partitionings on problem solving episodes. Second, the details of problem solving, as problem solving relates to the content and organization of the knowledge base, are more likely to be important than the general form of problem solving. And finally, expertise and its constituents must be studied in problem domains where subjects have exercised practice and have adapted - using tasks reasonably similar to those they normally encounter. Expertise has little transfer; it is the grindstone to which the expert has had his nose that counts. These implications from non-medically oriented literature are further supported from research on medical diagnosis.

Research on Medical Diagnosis

Recent research in clinical diagnosis (Barrows, Feightner, Neufeld, & Norman, 1978 ; Elstein, Shulman, & Sprafka, 1978; Kassirer & Gorry, 1978; McGuire &

Bashook, 1978) has contributed to a growing consensus about the general form of the process of diagnostic reasoning. Cues in patient data (signs, symptoms, laboratory test findings, etc.) suggest diagnostic hypotheses which are, in turn, tested against subsequent data of the case. This basic hypothetico-deductive process is shared by experienced and inexperienced diagnosticians alike, as are numerous parametric characteristics of the process such as the percentage of data items to first hypotheses, the average number of hypotheses maintained in active consideration, etc.

Despite their prevalent findings of lack of differences in the form of diagnosis as a function of experience, these research efforts have pointed the way to where important differences may lie and again, it is the knowledge base that is implicated. The Michigan State group (Elstein et. al., 1978) found that, with experience, physicians differ in the "accuracy of interpretation" of patient data with respect to the hypotheses they consider; experienced physicians are more likely to interpret findings appropriately as positive or negative evidence for the existence of a disease. This finding would appear to implicate the importance of knowledge of patient data that present in patients with particular diseases. The group at McMaster (Barrows

et. al., 1978) have found that experience can be discriminated by the actual hypotheses (as opposed to number etc.) that physicians consider during the diagnosis of a case. This suggests that experienced and less experienced physicians differ in their knowledge store of diseases or the cues by which they judge that particular diseases are likely to apply to a case. The same projects have also affirmed the problem-specificity of skill in diagnosis. The same physician may show different profiles of competence depending on his particular experiential history with different types of cases, a further indication that diagnostic skill, like other cognitive skills, is not entirely general, but rather is strongly dependent on the contents of problem-related knowledge.

Research at the University of Minnesota has concentrated on diagnosis in the medical subspecialty of pediatric cardiology and has resulted in a theory of diagnosis in this field that attempts to explicate knowledge and knowledge organization necessary for expert diagnostic performance (Johnson, Feltovich, Moller, & Swanson, 1979). Extensive experimentation and consultation with an expert pediatric cardiologist has resulted in a computer-runnable instantiation of the theory for this subject that represents knowledge explicitly

and shows strong correspondence to the subject's performance over a broad range of cases (Swanson, 1978; Swanson, Feltovich, Johnson & Moller, 1979). The documentation of this expert subject's knowledge-base yields some guidance regarding the content and organization of medical knowledge in the highly experienced diagnostician. However, the knowledge base of less experienced individuals has not yet been studied.

The Present Study

Within the constructs of the Minnesota theory, the present research attempts to assess differences in the medical knowledge base of individuals with increasing degrees of training and clinical experience in pediatric cardiology and the consequences of these knowledge differences for diagnostic performance. These cross-sectional differences provide evidence of how the knowledge base changes and develops as individuals progress from noviceness to expertise. In this section, "disease knowledge", the particular knowledge-base construct to be investigated, is discussed first. Some speculations about the nature of developmental differences in the disease knowledge base are then presented which provide guidance for the design of the diagnostic tasks used in the study to assess disease knowledge and

its diagnostic implications.

The particular knowledge base construct of focus in the present work is "disease knowledge."¹ Disease knowledge refers to a memory store of disease models. Each disease model is a memory structure that represents a disease. The disease model includes the pathophysiology of the disease and the signs, symptoms, and other clinical findings to be expected in a patient who has the disease. The model represents the physician's physiologic and clinical "picture" of the disease and can be thought of as a "schema" or "frame" as these constructs are used in cognitive psychology (e.g., Minsky, 1975; Rumelhart and Ortony, 1977). Constructs similar to the disease model have also appeared in other cognitive theories of medical diagnosis, for example, Rubin's (1975) "disease templates" and Pople's (1977) "disease entities."

1

The term "disease knowledge" refers to the same construct that in earlier expositions of the Minnesota theory has been termed "prototype knowledge" (e.g., Johnson et. al., 1979a). It was decided to abandon this latter designation because of its suggestion of entities typical of a class (e.g., Rosch and Mervis, 1975). While some disease models are prototypic, not all of them are.

In the theory of the expert, the set of disease models is extensive (see also de Groot, 1965; Simon & Chase, 1973) and organized hierarchically in groups and subgroups (see also Wortman, 1972; Pople, 1977). At upper (more general) levels of the hierarchy are disease categories, sets of diseases that present similarly because of physiologic or clinical similarity. Particular diseases occupy middle ranks of the hierarchy and these, in turn, are differentiated at the lowest hierarchical levels into numerous specific variants of each disease. Each disease variant may present differently in the clinic for reasons of subtle underlying difference in pathoanatomy, severity, or age of presentation in a patient.

While our previous work provides some guidance about the nature of disease knowledge in the expert, little is known about the knowledge base of the novice and how it changes toward expertise. Speculations about characteristics of the novice's disease knowledge can be garnered from analysis of the training experiences that novices encounter, the training materials they use, as well as psychological theory pertaining more generally to the development of knowledge bases. The first postulate for the novice's store of diseases is that it is classically centered. Initial training materials

(e.g., Moller, 1978) as well as the probability distribution of diseases presenting in the hospital, accentuate the most common diseases and the most common versions of diseases. These "training diseases" constitute "anchorage points" for subsequent elaboration of the store of diseases (see also, Ausubel, 1964; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1978, "basic objects"). A second postulate for novices is that the disease store is sparse in the sense that it lacks extensive cross-referencing and connection among the diseases in memory (Chi et.al., 1980; Elstein, Loupe, & Erdmann, 1971; Shavelson, 1972; Thro, 1980). It is with experience that the starting-point store of diseases is augmented, and both generalized into categorical clusters, as similarities among diseases are discovered, and discriminated into finer distinct entities as differentiation points among and within diseases are learned (Reed, 1978 ; Rosch & Mervis, 1975; Wortman & Greenberg, 1971). A third postulate about the novice disease store refers to the internal structure of the disease models themselves; this involves imprecision in the patient findings (signs, symptoms, etc.) to be expected clinically in a disease. Given that there is a range of natural variability associated with the clinical findings that can occur with any disease, large

sampling, through clinical experience or other training devices, is probably necessary to "tune" (Anderson, Kline & Beasley, 1979; Rumelhart & Norman, 1977) clinical expectations in disease models to the naturally occurring range. Novice expectations may be either overly general, tolerating clinical findings in a patient that should not occur, or overly specific, not allowing the legitimate range.

In contrast to the novice, whose disease store is assumed sparse, imprecise, and classical, the expert store of diseases, by converse arguments, as well as our prior research findings, is speculated to be dense, precise, and penumbral. "Density" refers to extensive cross-referencing and interconnection among diseases in memory. This is partly achieved through the coexistence of diseases within categorical groupings. A "penumbral" memory store of diseases includes less common and less stereotypic diseases and disease variants in addition to those included in the more standard "training set". "Precision" refers to clinical expectations within disease models which are tuned to be neither too tolerant nor overly restrictive.

The device for studying these knowledge claims in the present experimental study is the careful selection for diagnosis by subjects of naturally occurring patient cases each of which, through the structure of patient data it contains, provides a focused test of a different aspect of disease knowledge. An attempt was made to create diagnostic problems capable of showing where a diagnostician lies on various dimensions of knowledge base development. In a laboratory setting, these cases were diagnosed by subjects at different levels of experience with pediatric cardiology.

2. METHOD

Materials

Stimulus materials for the study were sets of patient data, each representing a different patient case, extracted from medical records of clinical cases seen at the University of Minnesota Hospitals. Clinical findings from the medical record for each case were assembled in a typed "patient file." The file arranged these data in the typical clinical order of history findings, followed by those from physical examination, x-ray, and electrocardiogram.

Cases. Five cases were used in the present study, each of which was chosen to assess a different characteristic of subjects' disease knowledge, for example, the differentiation of a disease into subtypes. In addition, the case design employed a "garden -path" methodology in that some chosen cases showed an early strong cue for an erroneous disease but had later critical, disconfirmatory evidence for this same disease. This device had three important functions. First, it enabled bringing all subjects to a common, comparable point in their thinking about possible explanations (hypotheses) for a case. Secondly, the garden-path allowed assessment of the precision in a subject's

model for the originally induced disease. For a subject to appropriately reject this disease, his correct evaluation of certain patient data items would be germane. Thirdly, because the true disease in a case was physiologically and clinically similar to the initially induced disease (a kind of "foil"), the garden-path established an environment for assessing the diseases that subjects considered as plausible competitors to the initial (foil) disease. The logical and temporal relationship between a subject's hypothesis generation of the initial disease and other reasonable candidates (including the true disease) could be studied.

The rationale which guided the selection of cases to be included in the study will now be discussed. Assessing the precision in subjects' disease models was a general objective that was to be addressed in all cases to be chosen - where precision refers to the accuracy of interpretation of patient data items as confirmatory or disconfirmatory evidence for a disease. For subjects to be compared on their precision, it would be helpful if the case caused all subjects to generate or consider a particular disease hypothesis in common during the course of the case. It was also useful for this shared hypothesis to be wrong,

forcing successful diagnosticians to interpret correctly the discrepancy between the hypothesis and certain data items.

This strategy would fail in a case if, in fact, the "foil" were not generated as a hypothesis by most, if not all, subjects in the study - from the least to the most experienced. Given our presumptions about the "classicality" of disease knowledge in inexperienced subjects, it was judged that the "foil" should be a common or "classic" disease to ensure that even the least experienced subjects would create it as a hypothesis.² In addition, such a "foil" would enable us to address another objective of the study which was to assess the relative "classical dependency" of subjects with different degrees of experience. From

² While a formal definition of a "classic" disease or disease variant cannot be given, the "spirit" of this designation can be discussed. Classic diseases or variants are ones that have a high clinical incidence (i.e., are relatively common) among patients with congenital heart disease, and/or receive the bulk of emphasis in introductory training materials. For example, in the introductory textbook used by subjects, Valvular Aortic Stenosis occupies five pages while Subvalvular Aortic Stenosis has one and one-half pages and Supravalvular Aortic Stenosis has two (Moller, 1978, pp. 95-103). In addition, Valvular Aortic Stenosis is the first of the three presented.

the above arguments, a selected case was to have an early strong cue for a "classic" disease but yet contain subsequent data that could fairly unequivocally prove this disease wrong.³ (For particular reasons, some cases deviated slightly from this general form and these deviations will be discussed under the descriptions of the particular cases in the study)

If cues were to exist in a case such that the classic foil and the true disease would, in fact, both be strong candidates, at least for a time, then these diseases would almost necessarily be structurally (anatomically and physiologically) similar, and, hence, constitute good candidates for existing in some subjects' memories as members of a disease category or grouping. The existence of various memory groupings ("upper" and "bottom" levels of the hierarchical structure of disease knowledge discussed earlier) in

³ The "strength of a cue (patient finding) for a disease is related to how reliably the cue is produced by the disease and to the number of other diseases that also produce the cue. Hence, a "strong" cue for a disease is one that presents reliably in patients with this disease and does not present in many other diseases (See Beach, 1964, on "cue validity" or any exposition of Bayes' theorem). While cue strength was not treated formally in the design of cases, the general guideline given here was sufficient for our purposes.

subjects could be tested by examining subjects' use of other diseases structurally, and hence, clinically similar to both the "foil" and true disease. The method for defining these target disease groupings will be presented later under the section on data and analysis.

The actual selection of cases was carried out with considerable help and guidance from an expert pediatric cardiologist and consultant to the project.⁴ The consultant was first approached with a general rationale for each case, for example:

A case where the patient's true condition is an uncommon variant of a disease and where there are early cues for a more classic variant of the same disease, but later data are clearly discrepant with this classic interpretation.

The consultant often had immediate hunches as to the kind of case that would fit the description; at other times he would think about the problem for a number of days. In either event, the consultant eventually produced a set of candidate cases from the Medical School patient files. The author and consultant ultimately

⁴

This person is on the faculty of the Division of Pediatric Cardiology and has been a collaborator in the study of diagnosis in congenital heart disease since the inception of the project at Minnesota.

chose from among these candidates after discussions, medical literature investigations, and trial diagnoses on the part of the consultant.

Synopses for the five cases selected are given below:

Case 1. The operative (true) disease in this case is Subvalvular Aortic Stenosis, an uncommon variant of Aortic Stenosis, the "classic" or most common version of which is Valvular Aortic Stenosis. The case is meant to assess subjects' differentiation of diseases into subtypes and the precision in their models of the classical variant..

Case 2. The operative disease in this case is Total Anomalous Pulmonary Venous Connection (TAPVC). The case contains classic auscultatory findings for Atrial Septal Defect (other findings are discrepant), a highly common congenital heart disease, findings that are also perfectly consistent with TAPVC, and, in fact, also consistent with any disease in the category of diseases with "volume overload in the right side of the heart." (Including, in addition to diseases mentioned, Partial Anomalous Pulmonary Venous Connection and some forms of Endocardial Cushion Defect). The case is designed to assess subjects' knowledge of and use of disease clusters corresponding to

disease categories, along with the standard precision issues.

Case 3. This case is a straightforward presentation of the operative disease, Patient Ductus Arteriosus, a highly common congenital heart disease. The case is intended to assess the relationships of this disease to other diseases in the subject's disease store and the diagnostic use of these related diseases in a case where the true diagnosis seems clear. This case was included so that there would be a case in the study which involved no "foil."

Case 4. The operative disease in this case is Pulmonary Atresia, one of a group of physiologically similar diseases (including, in addition, Tricuspid Atresia and Ebstein's Malformation) that constitute a category of "cyanotic diseases with decreased pulmonary blood flow."⁵ Like Case 2, this case is designed to assess subjects' knowledge and use of disease clusters corresponding to categories.

Case 5. The true condition in the case involves "multiple" diseases. The case contains an early

⁵ The medical reader may wonder about the absence of Tetralogy of Fallot from this group. Since it, unlike the other diseases in the group, has its shunt at the ventricular level, it is not so clearly a member of this group as are the other diseases included - for the purposes of this study.

confirmatory cue for the disease, Coarctation of the Aorta. Later data are discrepant with this disease in isolation and require not a shift to a different disease nor disease variant, but rather, the inclusion of additional diseases in a conjoint diagnosis. The original objective of this case was as a test of precision in subjects' disease models for Coarctation.

The taxing procedure for choosing cases should not be interpreted as meaning that the cases involved in the study are somehow "freaks". Although the true diseases, on the whole, are not among the most common congenital heart diseases, neither are they rare (e.g., all, except the Case 5 combination, receive at least some discussion in a standard introductory textbook of the field; Moller, 1978) and the clinical data used in the study to represent the diseases are not atypical. The fact is that all cases of TAPVC will, in some sense, have ASD as a "foil", and so forth for the other cases. The procedure for selecting cases was complex because of our attempt to find the clearest and best case to address each issue of interest.

Patient Files. The patient files to be presented to subjects were assembled after the choice of patient cases was made. The "patient file" is a

listing of the clinical signs and symptoms and the laboratory test results (from x-ray and EKG) that were observed in a real patient and recorded in the patient's medical record by the original attending physicians. Some of the guidelines used in assembling the patient file will be discussed first, followed by the actual procedure that was used.

An objective of the study was to be able to compare the inferences, interpretations, and evaluations of subjects in a uniform stimulus environment. For this reason, the order and content of patient findings to be presented to subjects was fixed. While this eliminated some components of the diagnostic process, primarily those associated with data collection and first-order interpretations of patient data (e.g., reading x-rays), "fixing" of the stimulus was crucial to the control needed to study the knowledge-based issues of interest. Because we wanted the variability among subjects to be primarily a function of their disease knowledge, attempts were also made to eliminate extraneous sources of variability arising from the patient data themselves, for example, ambiguity, lack of clarity, or conspicuous omissions. The intent was to make the statements of patient findings

as high quality as possible, consistent with a case of a particular kind. (The inherent quality of findings that can be achieved in some types of diseases and patients, e.g., newborns, is lower than in others) To this end, several precautions were taken in assembling information from the medical records.

First, cases were chosen for inclusion in the study only if the original attending physician was an established staff pediatric cardiologist (as opposed to, say, an intern or resident). This was an attempt to ensure the quality of the history and physical examination data. Secondly, the x-ray and EKG findings for a case were taken verbatim from the "formal reports" of these lab tests included in the medical record. The "formal report" is the hospital's official interpretation of an x-ray or EKG. Formal reports are created only by a select group of x-ray or EKG readers. Finally, minor modifications and additions were made to the findings as stated in the medical file if, in the judgment of the author and consultant, these were necessary to eliminate non-task-relevant sources of confusion. More will be said of the nature of these modifications below.

In assembling a "patient file", the consultant reviewed the medical record and extracted the first (verbatim) listing of patient findings. This first step involved a level of screening. There are some findings in a medical record that are routinely collected on a patient but are not generally important for the diagnostic work of a pediatric cardiologist. These include things like "financial history", "social history" (e.g., parent occupation, number of bedrooms in the house), and routine physical examination of HEENT (head, eyes, ears, nose and throat). Items of this type were included only if the consultant felt they were important in a particular case. The general guideline the consultant followed in choosing information was to include those items that he would send to another pediatric cardiologist as a description of a case. The types of findings that were included are given in Appendix A.

The first listing of findings was then ordered and segmented into small, meaningful units, each unit containing information on a different topic. The grossest level of ordering, History, Physical Exam, x-ray, and EKG was not problematic since it conforms to standard clinical procedure and, in fact, the medical files

are arranged this way (x-ray and EKG may be reversed). The ordering of findings within history and physical was determined by the consultant who chose an order he judged would be reasonably consistent with the order of information-gathering most practitioners in his field would use. Ordering was not an issue within x-ray and EKG as the formal reports were short and were used intact.

Within history and physical, data were segmented into small groups or units that could be presented to subjects, one unit at a time. Each unit might contain a few signs and symptoms, but the characteristic of a unit is that all of its information is about a logically coherent topic. For example, "first heart sounds" and "systolic murmurs" were included together because both give information pertaining to the same segment of the cardiac cycle and a diagnostician would be handicapped in analyzing this cycle component without both sources of information. Appendix A represents a good guide to the general segmentation structure built into the cases.

Data were segmented into small presentation units for two reasons. First, it encouraged fairly uniform, as opposed to sporadic, responding by subjects

over the course of a case. Secondly, it provides focus for determining the functional stimulus for particular subject responses (e.g., creation of hypotheses or interpretations of findings) during the diagnostic process.

After patient findings were ordered and segmented, all findings were reviewed by the author and consultant together for their clarity, precision, and completeness. The intent here was not to change the absolute character of the findings as the attending physician originally reported them, but, rather, to eliminate any troublesome artifacts that may have entered the findings through the mechanics, conventions, and implicit assumptions involved in transforming what was observed in the clinic into a medical record report.

As a result of this review, two primary types of modifications were made to the patient data. First, in some instances, data were added to the case. Data may be excluded from a medical record sometimes simply because of negligent reporting procedure but, more commonly, because of a tendency for physicians to report what they consider to be the "eventful" findings and to assume that the reader of the report will infer unreported findings to have been unimportant (e.g., "normal").

If some item in our "template" of findings for a case (Appendix A) did not appear in the medical report, the consultant fabricated a finding consistent with the other findings in the case and with the true diagnosis (examples of data sometimes added were birth weights and pregnancy histories). The attempt was to anticipate situations where a subject might feel that "if he only knew x", he could clarify some issue. The second type of modification to data items involved minor embellishments or clarifications to the statements of particular findings. Again, most of these were necessary because of the implicit assumptions of those who created the medical reports. For example, if a physician hears a very soft murmur he may not say anything in the report about the direction of radiation of the murmur because soft murmurs do not radiate much. It was our decision not to leave such inferences to chance but rather to state them explicitly, e.g., "There is little radiation from the murmur."

No addition or change to the findings as reported in the medical record was made if, in the judgment of the consultant, some suboptimal component of the record was due, not to some aspect of

reporting, but, rather, to some inherent characteristic of the case. Characteristics of patients and particular diseases set constraints on how "good" the findings can be. For example, precise findings are hard to get from newborn patients in general. Hence, in Case 4, Pulmonary Atresia in a 4-day old child, the murmur descriptions are not as precise as some subjects would have liked; however, the consultant's judgment was that they are probably as well specified as they could have been in the clinic for a child of this sort.

The final content, segmentation structure, and ordering for findings in each of the "patient files" representing the five cases used in the study are given as Appendix B.

Subjects

Subjects were twelve volunteer (i.e., unpaid) individuals from the University of Minnesota Medical School who were chosen to span a dimension of clinical experience in the diagnosis and management of congenital heart disease. Two were female (both "trainees") and the remainder male. There were four subjects from each of the following three groups:

Students. These were fourth-year medical students who had just completed a six-week course in pediatric cardiology. This course has a five-day per week classroom component in which most congenital heart diseases are covered, although not all are given the same emphasis nor commitment of time (see Moller, 1978). As part of this training, each student held primary responsibility (with supervision) for diagnosis and management of 25-50 patients with congenital heart disease.

Trainees. Subjects in this group were either in the third year of a general pediatrics residency or were beginning their first year of fellowship in pediatric cardiology. General pediatrics residents are individuals who have completed medical school and, in their residency, acquire considerable clinical

experience in the diagnosis and management of children with a broad range of disorders, including disorders of congenital heart disease. The residency usually lasts three years. Residents used in the study had just completed a rotation in pediatric cardiology. Fellows are individuals who have completed a pediatrics residency. Their training during fellowship is focused on the diagnosis and management of patients with congenital heart disease. The fellowship is the academic experience designed to train practitioners of the pediatric cardiology subspecialty. Subjects in this group estimated that they had held primary responsibility for about 150 patients with congenital heart disease. Residents and fellows did not differ in their estimates.

Experts. This group was composed of two faculty members in the Division of Pediatric Cardiology with upwards of twenty years of practice as pediatric cardiologists and two fourth-year fellows in pediatric cardiology. One fellow had just become Board certified as a pediatric cardiologist and was appointed as an instructor in the University of Minnesota Medical School.

The two fellows estimated that they had held primary responsibility for about 400 patients with congenital heart disease. The best estimates the faculty subjects could give were somewhere between five and ten thousand. The experience discrepancy within the "expert" group was intentional. We wanted to be able to compare subjects of extreme levels of experience (faculty) with those who, although prepared to practice the subspecialty, lacked such extensive experience (fellows). It was hoped that this comparison would yield insight into how the knowledge base is "fine-tuned" after it is basically established.

Sampling of subjects within the experience levels was not a functional issue. The Division of Pediatric Cardiology is a very small subspecialty unit of the Medical School. Except for the two faculty subjects, other subjects represented nearly all people who existed at the experience levels of interest. In fact, this was the reason that both residents and first-year fellows had to be combined to form a "middle-experience" group. Additional individuals existed at the faculty level. The two who were chosen were selected because of availability, willingness, and

because the consultant considered them to be outstanding.

The relatively small "sample-size" (12) was necessary partly because of the small subject pool and partly because of the nature of the research. The research is idiographic in nature in the sense that it requires anticipation and analysis of each subject's knowledge base as this interacts with a carefully analyzed and structured information environment. The preparation of materials, the time commitments required from subjects to carry out the tasks (3-4 hours per subject), as well as the analyses of data required to do such work set severe limits on the number of subjects that can be handled.

Subjects used in the study were contacted and recruited in the following manner. The consultant provided the author with a list of names of people at the prespecified experience levels. At the same time, the consultant sent a note to each potential subject, introducing the author and informing the candidate that the author would be calling to solicit his or her help. All potential subjects subsequently contacted by the author agreed to participate without remuneration.

Procedure

All subjects participated during a one-month period. Two two-hour sessions were scheduled for each subject. The length of time between sessions varied across subjects from one day to a week, depending upon the subject's work or school schedule. Subjects were asked not to discuss the study or cases with anybody else until after the month in which the experiment was conducted.

Each subject diagnosed all five cases and every subject diagnosed the cases in the same order. Diagnosis of each case required approximately one-half hour. During the first session, the subject worked cases one (Patent Ductus Arteriosus), four (Pulmonary Atresia), and three, (Total Anomalous Pulmonary Venous Connection), in that order. In the second session, subjects diagnosed in order, case five (Coarctation-plus) and case one (Subvalvular Aortic Stenosis). No discussion of cases or a subject's performance was conducted with subjects and, in particular, the correct "answers" (diagnoses) were not given to subjects until the end of the second session, after all cases had been diagnosed. After the final case, there was an informal "debriefing" session during which the correct

diagnoses for all of the cases and the subjects' performance and impressions of the cases were discussed.⁶ Debriefing sessions varied greatly in length, from thirty minutes to four hours, depending on the subject's willingness and time commitments. Where time permitted, an attempt was made to informally "quiz" subjects about various aspects of a case (e.g., "Can you discuss the EKG Axis in Pulmonary Atresia, Tricuspid Atresia, and Ebstein's Malformation?") before correct diagnoses were given. Information from these debriefing sessions was not used in any formal way in the study, but did, at times, contribute to an understanding of subjects' performance on a case.

A quiet, comfortable, private office was provided by the Division of Pediatric Cardiology for conducting the sessions. Subjects were given the option of using this office or any other place they wished, provided that it was quiet and sessions would not be

⁶

Except for Case 5, the correct diagnosis for a case was established by cardiac catheterization, an invasive probing procedure of the heart which is usually definitive for diagnosis. The exception for Case 5 will be discussed under the "results" section for that case.

interrupted. Alternative settings used ranged from professional offices to private homes.

The subject was seated at a desk or table. The experimenter (the author) sat facing in the same direction as the subject, off to one side, and slightly behind the subject. This was so that the experimenter would not be in the subject's view as the subject worked.

The subject himself, first read the instructions for the task (Appendix C). The experimenter then read the instructions aloud. (Second sessions involved only the silent reading). Following clarifications requested by the subject, the subject was presented the "patient file" for a case. Instructions directed the subject to read aloud each numbered data segment (see Appendix B) in the order in which data were given in the patient file and to report aloud any thoughts he had toward formulating a diagnosis for the patient's condition. If he had no thoughts following an item, he was to pass on to the next. The subject was free to review any item he had already seen, but was asked not to skip ahead. When reviewing, the subject was asked to re-read the item aloud so that we would know what he was attending to.

At four points in the case, after history, physical examination, x-ray, and EKG, the subject was asked for an explicit reporting of any "hunches" he might have about the patient's condition. This involved a short, standard interjection from the experimenter: e.g., "Hunches after history please." At the end of the case, the subject was also asked for a primary diagnosis and as many as two alternatives. Except for asking for "hunches" and the final diagnoses, the only other interaction between the experimenter and a subject, once a case had begun, was an occasional request to "report your thoughts" when it was clear that a subject was thinking but not talking.

Sessions were tape recorded for later transcription.

Data and Analysis

Basic data from the study were typed transcriptions (protocols) of tape recordings made while subjects diagnosed the cases and reported aloud their thinking toward a diagnosis for each case.

Particular analyses of these data vary somewhat according to the objectives of each case. In general, analyses are organized around a concept of "Logical Competitor Sets" (LCS) which are sets of diseases targeted as important from the choice of cases for the study (see "Materials" above). Diseases in the competitor set for most cases share underlying physiology with the operative or true disease in the case and, hence, have similar clinical presentation.

In concentrating analyses on the Logical Competitor Sets for each case, a commitment was made to focus analyses on diseases specified in advance to be plausible alternatives for the case and that are likely to be more difficult to discriminate among themselves than are diseases within this group with those outside it. Hence, they constitute a set of "good" hypotheses to be considered in a case. One major motivation for restricting analyses in this manner comes from prior work on expert-novice differences which

suggests that unless a dimension of quality is built into the "dependent variables" considered, expert-novice differences are not likely to be revealed (Chase & Simon, 1973a; Barrows et. al., 1978).

Another motivation is the case design itself for the study. Although it was assumed that disease hypotheses outside the LCS for each case could be considered by subjects at various times, it was anticipated that the structure designed into the stimuli (case materials) would greatly control and restrict subjects' performance and, most importantly, that the important dynamics of each case would center around the prespecified hypotheses (the LCS) and their management.

The LCS for each case was developed from two major sources. First, for the operative (true) disease in each case, the expert in pediatric cardiology and collaborator on the project, was asked to specify the set of alternative diseases most similar to the true disease and likely to be confused with it. Because these are diseases that are highly similar in clinical presentation, he was also asked to specify items of patient data which, if interpreted correctly, could be used to discriminate among diseases in the LCS. These

judgments were then cross-checked against a major disease reference for pediatric cardiology (Moss, Adams & Emmanouilides, 1977). Specifically, for each disease described in this reference, the authors provide a "differential diagnosis" section which discusses diseases similar to and difficult to discriminate from the target disease, as well as differential data points. Based on the reference, no diseases were deleted from the consultant's list although some were added.

For each case, protocols were coded for the two general kinds of uses of the Logical Competitor Set. The first of these is the use of LCS members as hypotheses by subjects at each patient data point of the case. To the extent LCS members are used together, this is taken as evidence that these diseases are being used as competitors and are clustered in memory. The second is the evaluations of LCS members with respect to a set of selected data items. These evaluations yield evidence of the precision in subjects' individual disease models and also can be used to discern characteristic kinds of errors among the subjects and the loci of these errors in disease knowledge.

All tables and interpretations of subject protocol reported in the body of this paper are based on the coding and judgments of the author, a person who has worked in the subject matter of pediatric cardiology and with data of this sort for over five years. The author believes such knowledge and experience contributes greatly to the understanding of subjects' overt behavior. These codings and interpretations were "blind" neither to the identities of subjects nor to the objectives of the study.

Recently the laboratory which sponsored the present study has conducted an investigation of coding reliability on the protocols generated from the present study. The protocol coding addressed in that investigation was of the extraction of all hypotheses used in a case by subjects which include as a subset the hypotheses of interest in the present study, that is, the Logical Competitor Set members. It was found that the hypotheses used by subjects could be extracted with respectable reliability by multiple scorers (See Appendix D for a description of this investigation and its results).

In the present study, the major protocol coding for one case, Case 1, was submitted to an alternative scorer to establish a degree of agreement between a

second person and the author. The coding involved the identification of LCS members used by subjects during the course of the case and evaluations by subjects of target data items. The second scorer was a senior college student with some interest in medicine and some knowledge of pediatric cardiology. The coding by this subject was "blind" to the identity of subjects. The general scoring procedure used by the second scorer is given as Appendix E. More detailed accounts of this alternative scoring and discussions of the agreement between the author and the second scorer will be introduced at appropriate points under the "results" section for Case 1.

3. RESULTS

In this section, the results of the study will be presented in a case by case manner. The presentation of results for each case will follow the same general format. First, there is a brief description of the objectives of the case in terms of the knowledge base issue it is intended to address. The disease representing the operative disease in the case (the true diagnosis) is then discussed along with other diseases in the Logical Competitor Set for this condition. The basic structural properties (anatomy, physiology) of these diseases are presented along with their important clinical findings, especially the patient data items which can potentially be used to discriminate among the LCS members. These differential data items were the ones considered in advance to be the case findings on which subjects' diagnoses of the case would "turn"; that is, it was anticipated that subjects' handling of these items would have much to do with their successful or unsuccessful diagnoses of the case.

Two kinds of results are then presented for each case. The first of these, the "Use of Logical Competitor Set", presents the members of the Logical

Competitor Set that were used by subjects as hypotheses at each patient data point of the case. To the extent LCS members were used by a subject as diagnostic hypotheses and were used in close proximity to the important cues for any one LCS member, this is taken as evidence that the LCS members exist within memory in some form of interconnected cognitive unit (Anderson, 1980), for example, a category. This bears on the issue of "density" in disease knowledge. In those cases that have a classic "foil", the use as hypotheses of the other, less classic, LCS members in addition to the foil, bears on the issue of an extended or "penumbral" disease knowledge base versus one that is more classically dependent. The second type of result, "Diagnostic Errors", is an analysis of errors subjects made in diagnosis of the case. This is a kind of "bug" analysis (Brown & Burton, 1978) in that it attempts to identify sources of misconception in subjects' knowledge responsible for error. An attempt is made to explain the performance of each subject misdiagnosing the case by identifying the critical issue or issues of the case that led the subject astray and, in addition, to identify characteristic commonalities of error and disease knowledge deficiency among groups of subjects.

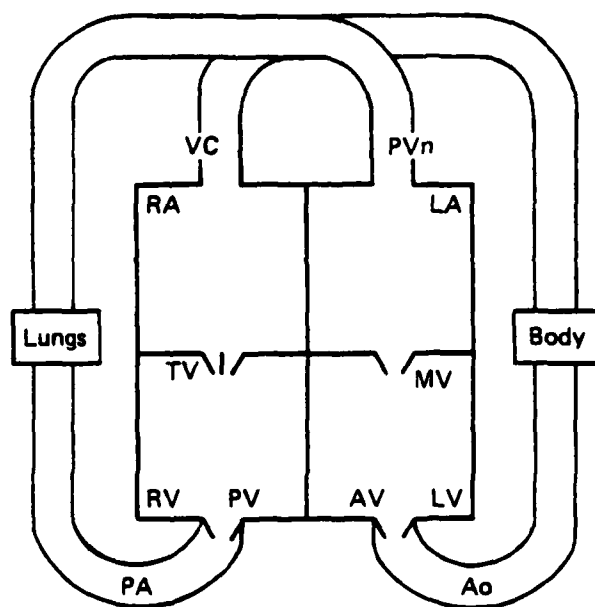
In many instances, critical errors pertain to subjects' evaluations of prespecified patient data items in relation to LCS members; however, in those instances where the critical error is outside the prespecified data item or diseases, other items are introduced. In large part, error analyses provide evidence for types of "imprecision", within subjects' disease models, associated with clinical expectations for a disease. However, other sources of error are also identified.

Because the presentation and discussion of results within this section make extensive reference to the heart and cardiovascular system, a very brief introduction to the anatomy and physiology of the cardiovascular system is given as a prelude to the section - in order to help the reader better understand what comes thereafter.

The Normal Cardiovascular System

Figure 1 shows the normal heart and other major components of the cardiovascular system.⁷ Starting on the right side of the heart, the right ventricle (RV) of the heart pumps blood across the pulmonary valve (PV), through the pulmonary artery (PA), and into the lungs where the blood receives oxygen. Blood then returns to the heart via the pulmonary veins (PVn) into the left atrium (LA). From the left atrium, oxygenated blood proceeds across the mitral valve (MV) into the left ventricle (LV), where it is pumped across the aortic valve (AV), through the aorta (Ao), and to the body. In the body, oxygen is extracted from the blood which then flows back to the right atrium (RA) of the heart via the vena cavae (VC). Deoxygenated blood from the right atrium flows across the tricuspid valve (TV) into the right ventricle and the cycle repeats. The

⁷ In Figure 1 and all figures like it presented in the report, "left" and "right" are from the "patient's" point of view, i.e., the reverse of the perspective of the reader viewing the figure. Also, all major anatomical components are given a symbol in Figure 1, and the names of components referenced by the symbols are given explicitly. Only the symbols for anatomy germane to the discussion will be used in subsequent figures; if need be, the reader can consult Figure 1 for their meanings.



LEGEND

Ao = Aorta	PV = Pulmonary Valve
AV = Aortic Valve	PVn = Pulmonary Veins
LA = Left Atrium	RA = Right Atrium
LV = Left Ventricle	RV = Right Ventricle
MV = Mitral Valve	TV = Tricuspid Valve
PA = Pulmonary Artery	VC = Vena Cavae

Figure 1. The normal heart and cardiovascular system.

"upper" chambers of the heart, the atria, are normally separated by the atrial septum, while the "lower" chambers, the ventricles, are normally separated by the ventricular septum.

Congenital heart diseases are anatomic or physiologic abnormalities within the heart and cardiovascular system (e.g., holes in heart septa, tight valves, or electrical conduction problems). These basic abnormalities alter the flow, pressure, or resistance patterns of the system and produce the patient manifestations (signs, symptoms, laboratory test results) that the physician must utilize in diagnosis. Particular diseases, pertinent to cases presented in this section, will be described under the statement of results for each case.

Case 1: Subvalvular Aortic Stenosis

The purpose of this case is to investigate subjects' differentiation of a disease into subtypes. The vehicle for doing this is a diagnostic problem which encourages subjects to display, in a diagnostic setting, their working knowledge of a set of disease variants.

The Logical Competitor Set for Case 1 includes three variants of Aortic Stenosis: Valvular Aortic Stenosis (ValvAS), Subvalvular Aortic Stenosis (SubAS) and Supravalvular Aortic Stenosis (SupAS). Figure 2 depicts the anatomical abnormalities within the heart which define each of these disease variants. All involve obstruction to left ventricular outflow with different variants defined by slight differences in the locus of obstruction: ValvAS is obstruction at the aortic valve itself; SubAS is an obstruction slightly "upstream" from the valve; SupAS is obstruction slightly "downstream" from the valve. Because these disease variants are only subtly different anatomically and physiologically, they differ only slightly in clinical presentation. ValvAS is the most common of the three and receives the greatest amount of exposition in introductory training materials of pediatric cardiology.

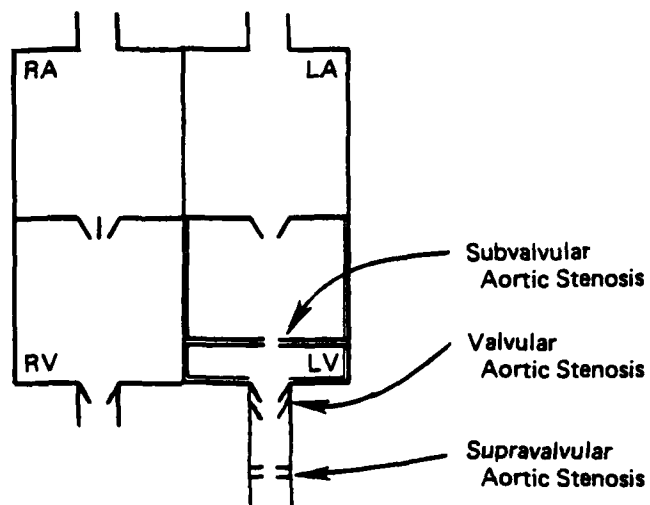


Figure 2. Logical competitor set for Case 1. Three types of Aortic Stenosis.

(e.g., Moller, 1978). Hence, it might be expected that subjects' knowledge for ValvAS will develop more rapidly than for the others and that ValvAS may function as a "foil" for some subjects. SubAS, however, is the operative disease in the case and the correct diagnosis.

In the "patient file" presented to subjects for Case 1 (Appendix B, Case 1), patient data items 17 and 19, a "thrill" and a "murmur" respectively, are strong cues for Valvular Aortic Stenosis although they are compatible with the other variants. Hence it was suspected that all subjects would raise at least ValvAS as a hypothesis by the time of these data points. Data item 18, a finding of "no systolic ejection click", is very strong evidence against ValvAS. Data items 10, "normal facies," and 22, "prominent aorta," are evidence against SupAS. All data of the case are compatible with the operative disease, SubAS.

Use of the Logical Competition Set - Case 1

Table 1 shows the variants of Aortic Stenosis that were used as hypotheses by individual subjects⁸ at all patient data points where any variant was

⁸ Subjects E3 and E4 are the faculty subjects with upwards of twenty years of experience. They are noted with an asterisk in this and all subsequent tables.

Table 1
Case 1: Subject Use of LCS Hypotheses in Response to Patient Data Items

Subjects/Hypotheses	Patient Data Items ^a																			
	History								Physical Exam											
	1	3	4	7	8	HHx	10	13	14	17	18	19	20	HPEx	22	HXray	23	HEKG		
S1 ValvAS			X	X		X					X	X		X	X	X		X		
SubAS							X													
S2 ValvAS				X		X				X	X	X	X	X	X	X		X		
SubAS																				
S3 ValvAS						X				X	X	X	X	X	X	X		X		
SubAS							X													
T1 ValvAS						X				X	X	X	X	X	X	X		X		
SubAS																				
T2 ValvAS																				
SubAS				X	X	X				X	X	X	X	X	X	X		X		
T3 ValvAS						X				X	X	X	X	X	X	X		X		
SubAS																				
T4 ValvAS										X	X	X	X	X	X	X		X		
E1 ValvAS	X	X		X		X				X	X	X	X	X	X	X		X		
SubAS																				
E2 ValvAS										X	X	X	X	X	X	X		X		
SubAS																				
E3* ValvAS						X				X	X	X	X	X	X	X		X		
SubAS																				
E4* ValvAS										X	X	X	X	X	X	X		X		
SubAS																				
SupAS																				

Note. (X) indicates a subject's use of a hypothesis at the time of a patient data item.
^a Statements of numbered patient data items are given in Appendix B, Case 1.
 * (HHx) etc. refer to points in the case where subjects are asked for hunches.
 * The two experts with more than twenty years of experience.

mentioned by any subject and at the four points of the case where "hunches" were actively solicited from the subjects (HHx, hunches after history; HPEx, hunches after physical; HXray, hunches after Xray; HEKG, hunches after EKG). Data items from the "patient file" (Appendix B, Case 1) are listed across the top in the left to right order in which they were presented to subjects. An "x" in this table simply indicates that the subject mentioned a particular Aortic Stenosis variant in his protocol at the data point where the "x" appears.

Instructions and guidelines for coding the variants of Aortic Stenosis used by subjects in protocol and for building Table 1 are given as Appendix F.⁹

⁹In reading Appendix F and Appendix H, the reader will note that in coding for Table 1 the author and alternate scorer coded not only for the simple mention of Aortic Stenosis variants at data points but, in addition, for the "tone" with which the variant was mentioned when it was mentioned, that is, whether it was spoken of as a "good" or "bad" explanation for the patient's condition. No use of these particular "tone" codings was made in the study and the reader need only attend to the aspects of Appendices F and H that bear on deciding whether or not a variant was mentioned at all. These appendices are included in the report as they are because the alternate scorer used them in this form in following the procedure for coding (Appendix E).

The coding of the alternate scorer and agreement with the author are addressed in Appendix G. An entire protocol for one subject, T3, is given as Appendix H along with a complete discussion of the author's coding of this subject for Table 1.

Table 1 shows an increase in the use of variants of Aortic Stenosis, other than ValvAS, from medical students to experts in pediatric cardiology. In particular, only one student, S2, ever raised both of the less classic variants of Aortic Stenosis at all, during the entire course of the case, and he mentioned SubAS and SupAS only once each. Two trainees (T1, T3) and three experts (E1, E3, E4) used all three variants at some time during the case. If one considers the number of subjects in each group who not only used all three variants, but used each more than once, no students, one trainee (T1) and, again, three experts meet this criterion.

While simple mention (as reflected in Table 1) of the Aortic Stenosis variants as hypotheses is one indication of whether these were considered by subjects, a measure of how actively these hypotheses were considered is the prevalence with which they were evaluated with respect to data items. Table 2 shows all

Table 2
Case 1: Evaluations of Target Data Items in Relation to LCS Hypotheses

Target Patient Data Items													
10		17			18		19		20		22		
Normal Facies		Thrill			No Click		Murmur		Aortic Insuff.		Prominent Aorta		
Hypotheses	ValvAS	SubAS	SupAS	ValvAS	SubAS	SupAS	ValvAS	SubAS	SupAS	ValvAS	SubAS	SupAS	
Subjects													
S1	-			-			+					+	
S2	-		+	+			+		+			+	
S3			+	+			+		+			+	
S4							+					+	
T1			+	-			+	+	+			+	
T2							+					+	
T3	0		+	-		+	+		+			+	
T4			+	-			+					+	
E1	-		+	0	+	+	+	0	+			-	
E2	-			-		+	+		+			-	
E3*	-		+	+		+	+	+	+			+	
E4*			+			+	+	+	+			+	

Note. (+) indicates subject judged data item as confirmatory for a hypothesis.
 (-) indicates subject judged data item as disconfirmatory for a hypothesis.
 (O) indicates subject judged data item as ambivalent in relation to a hypothesis.
 * The two experts with more than twenty years of experience.

evaluations by subjects of the Aortic Stenosis variants with respect to the set of data items which are central to successful solution of the case. A mark (+, -, o) under a disease variant and data item in this table indicates that the data item was judged to be positive, negative, or ambivalent evidence for the disease variant as a hypothesis.¹⁰ For example, a negative evaluation of NO CLICK with respect to ValvAS would be: "The lack of a systolic ejection click is against Valvular Aortic Stenosis." Instructions and guidelines for the coding in this table are given as Appendix I. The coding of

¹⁰

There is no absolute correspondence between the use of an hypothesis at the point of a particular data item (Table 1) and the evaluation of the hypothesis with respect to that data item (Table 2). Subjects can evaluate an hypothesis with respect to a data item long past (e.g., evaluate with respect to data item 10 when he has reached, say, data point 17 of the case) and can also mention an hypothesis at a data point without necessarily evaluating the hypothesis with respect to that data item. Hence, for example, even though subject S2 mentioned all three variants at data point 10, he only ever evaluated one of these (SupAS) with respect to data item 10. The mention of the other variants at 10 was as part of a puzzled attempt to recall the variants of Aortic Stenosis.

the alternative scorer on Table 2 and agreement with the author are addressed in Appendix J.

Table 2 shows an increase, from students to experts, in the active evaluation of data items as evidence for or against the variants of aortic stenosis. In particular, no student evaluated all three of the variants with respect to a data item (of course, only one student, S2, ever mentioned all three variants at all). The two trainees (T1, T3), and three experts (E1, E3, E4) who used all three variants in the case also evaluated all three variants with respect to at least one data item. While this suggests activeness in the evaluation of variants by more experienced subjects, it does not necessarily reflect comparative evaluation. However, when a subject evaluates all variants with respect to the same data item, this is an indication that the subject is actively attempting to weigh the variants against each other to determine which is the best explanation for the data item and case. In this regard, no students, the two trainees (T1, T3), and again, the three experts (E1, E3, E4) evaluated all three variants with respect to a common (the same) data item. These same experts, but not the trainees, evaluated all variants in relation to more than one data

item in common (E1, 5 items; E3, 2 items; E4, 2 items).

The analysis thus far suggests that with increasing diagnostic experience subjects know and actively utilize non-classical variants of a disease as hypotheses in a diagnostic setting. Examination of the two most experienced subjects, E3 and E4, yields some clue as to the knowledge structure that supports this performance. Figure 3 shows the protocols of these subjects at two data points, 17 which is the first strong evidence for Valvular Aortic Stenosis and other variants, and 18 which is the strongest evidence against ValvAS. E3 raises all three variants together (in the same "breath" so to speak) at the time of the first strong evidence. These hypotheses are then available to be evaluated comparatively against subsequent data, in particular, data item 18. This same form characterizes subject E1 (see Table 1). Subject E4, however, aggressively focuses on the "classic" member of the competitor set at 17, but immediately expands to the full set upon receiving strong negative evidence at 18. This form is shared by subject T3 and, less clearly, subject T1 (see Table 1).

One explanation for these patterns is that in the expert a disease and its set of subtle variations come

(17) There is a systolic thrill felt below the right clavicle, along the mid-left sternal border, and in the suprasternal notch.

E3: This thrill is most consistent with a diagnosis of bicuspid aortic valve or aortic valvular stenosis. It would also be consistent with supra-avalvular stenosis and discrete sub-aortic stenosis.

E4: Until proved otherwise, now, he must have valvular aortic stenosis.

(18) The first heart sound is normal and there is no systolic ejection click.

E3: The absence of a systolic ejection click in the presence of what I would consider to be an aortic outflow thrill makes aortic valvular stenosis and bicuspid aortic valve less likely. Aortic valvular stenosis of a very severe degree might be associated without a click. On the other hand, uh, it makes us think more seriously of discrete membranous subaortic stenosis.

E4: Absence of the click is against valvular aortic stenosis. Then perhaps instead he has subvalvular or supra-avalvular aortic stenosis.

Figure 3. Protocols from experts E3 and E4 at data points 17 and 18 in Case 1.

to constitute an interconnected memory unit, a kind of category; when one of the members is strongly activated in memory, the category and other members are also activated. The expert can then choose to consider category members in two modes. In the first mode, he tests all members simultaneously. This first mode might be termed "precautionary" since if any hypothesis encounters disconfirmatory evidence, alternative explanations for which the same evidence might be compatible are already under consideration. In the second mode, the expert tests only the most likely (in his current judgment) member. This mode might be termed one of "extraction" because its general success depends heavily on "rejection" of the target disease when appropriate which, in turn, depends heavily on the precision in the diagnostician's model for the disease. Once the target disease is rejected, other category members provide a ready "back-up" set of alternative hypotheses. Further evidence for these speculations will be addressed as results from other cases are presented.

Diagnostic Errors - Case 1

A final analysis of the results of this case involves an attempt to discern the causes for subjects' errors in final diagnosis. Table 3 gives the final

Table 3
Case 1: Subvalvular Aortic Stenosis — Final Diagnoses

Subjects		Final Diagnosis
Students	S1	Valvular Aortic Stenosis
	S2	Valvular Aortic Stenosis
	S3	Valvular Aortic Stenosis
	S4	Valvular Aortic Stenosis
Trainees	T1	Subvalvular Aortic Stenosis
	T2	Valvular Aortic Stenosis
	T3	Subvalvular Aortic Stenosis
	T4	Valvular Aortic Stenosis
Experts	E1	Valvular Aortic Stenosis
	E2	Valvular Aortic Stenosis
	E3*	Subvalvular Aortic Stenosis
	E4*	Subvalvular Aortic Stenosis

*The two experts with more than twenty years of experience.

primary diagnosis for each subject. Six subjects (S1, S3, S4, T2, T4, E2) never considered Subvalvular Aortic Stenosis at all (see Table 1) although all generated Valvular Aortic Stenosis. At least three explanations could apply to this lack of activation. First, and most basically, it could be that subjects do not know about SubAS at all. However, post-experimental interviews with all these subjects confirmed that they had some knowledge of this disease and could describe it. A second possible explanation is that these subjects have built up no strong "bottom-up" association in memory between any data item of the case and the subvalvular disease. Even lacking such a "trigger" for SubAS itself, it would have been possible for subjects to generate SubAS as a side effect of their activation of ValvAS, if these two diseases were related in a memory unit, through a process of "spreading activation" (Anderson, 1976) or "top-down" activation (Rumelhart & Ortony, 1977); Bobrow & Norman, 1975). This suggests the third explanation - that for these subjects, their knowledge representations for the variants of Aortic Stenosis exist more in isolation than they do in the more experienced subjects. This is the issue of sparseness in disease knowledge.

For those subjects who generated ValvAS as a hypothesis but failed to abandon it in the face of strong negative evidence, examination of their handling of this disconfirmatory evidence yields insight into the nature and precision of their disease models for ValvAS. Discussion will focus on data item 18, the strongest evidence against ValvAS.

Two students (S2, S3), evaluated 18, NO CLICK, as confirmatory for ValvAS (Table 2). This appears to reflect, simply, an error in important factual knowledge about this disease. Two subjects (S4, T2) did not evaluate 18 at all with respect to ValvAS (Table 2). Significantly, they also did not generate any variant of Aortic Stenosis until after data item 18 (Table 1). This suggests that the memory store of bottom-up associations between data items and Aortic Stenosis variants for these subjects is not as extensive as for other subjects and, in particular, that data item 17 is not recognized as a strong cue for Aortic Stenosis type diseases. A further implication is that the physical examination finding of a "systolic ejection click" and its import in ValvAS are not represented in the ValvAS disease models of these subjects since, if they were, the model itself should have lead the subjects to re-examine this finding.

Finally, there were four subjects (S1, T4, E1, E2) who, although evaluating 18 as negative for ValvAS, maintained ValvAS as a final diagnosis. The protocols of subjects S1 and T4 yield some insight into an explanation for these subjects. Figure 4 shows the protocols for these two subjects at data points 18 and 22, the latter consisting primarily of the finding of a "prominent aorta" on x-ray. Both subjects question ValvAS at 18, but are much more satisfied with this diagnosis at 22 and thereafter. Why might this be?

Figure 5 shows the causal relationship between a "tight" or stenotic aortic valve and an enlarged or prominent aorta. To open a tight valve, the left ventricle (LV) of the heart must generate abnormally high pressure. Blood expelled under this high pressure forces against the aortic wall and expands it. For the two subjects under discussion, it appears that their causal knowledge attributes the "systolic ejection click" in ValvAS to the enlarged aorta itself; that is, the click is caused by the large chamber into which the valve is opening, perhaps some kind of resonance phenomenon. For these subjects the causal chain from the valve to the "click" is as follows:

tight valve → big aorta → click

(18) The first heart sound is normal and there is no systolic ejection click.

S1: Ah, well this, the fact that there is no systolic ejection click present, tells us that there is probably not a post stenotic dilatation of the aorta which one would expect with the presence of aortic stenosis and some aortic insufficiency. However, this does not necessarily rule it out.

T4: Love it. Um, well, okay. I wonder if there is . . . , no click, that's funny. I would expect if he had AS. I wish they had said whether the murmur went up into his neck, Okay.

X-RAY

(22) The chest X-ray shows normal cardiac size and contour and normal vascularity, but prominence of the ascending aorta.

S1: Ah, well this is what one would expect with ah, aortic stenosis with secondary aortic insufficiency. One would expect that the aorta, ascending aorta distal to the ah, to the stenosis, would be dilated due to the changes in the wall tension across the gradient. Therefore, ah, the fact that ah, a click was not heard on physical exam, may have been a subjective finding of the person examining. But, the X-ray does indeed suggest that there is some post stenotic dilatation.

T4: Ha ha! AS-AI.

Figure 4. Protocols from subjects S1 and T4 at data points 18 and 22 in Case 1.

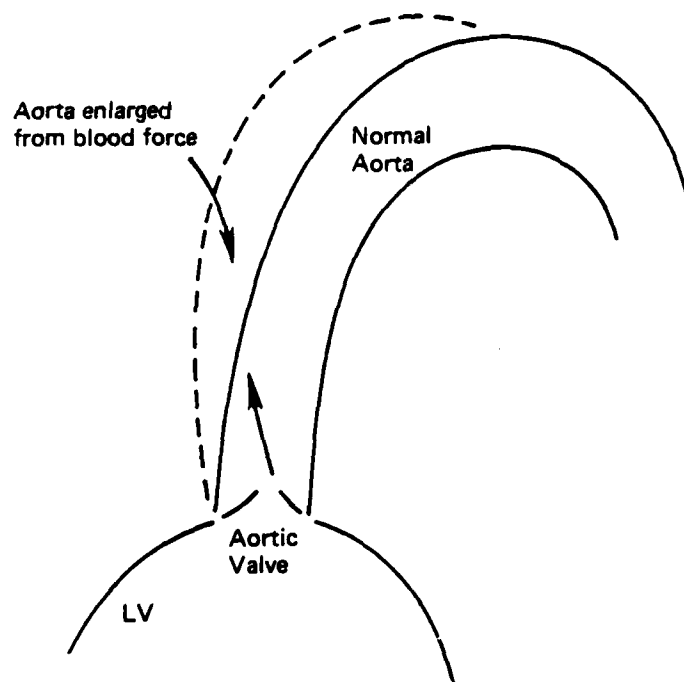


Figure 5. Aorta enlarged from the force of blood ejecting from a stenotic aortic valve.

Hence, for these subjects, the "big aorta" itself is predominant over the "click" as evidence for ValvAS, with the click just additional evidence for a big aorta. Once they receive their best evidence for a "big aorta", 22, they are no longer worried about the lack of a "click".

The true state of affairs appears to be that tight valve causes both the "click" and the enlarged aorta at the same level of cause (Friedman & Kirkpatrick, 1977, p. 180). The systolic ejection click is associated with the opening of the tight valve itself as shown below:

tight valve → click
 ↓
 big aorta

Hence, both of these effects must be proved. Why might a number of subjects have misconstrued this relationship? One need look no further than the introductory textbook these subjects use (Moller, 1978, p. 96) where the erroneous causal relationship is stated, or at least strongly implied.

The subjects just discussed raise two important issues. First, they demonstrate how "small" knowledge errors can have major repercussions for the handling of

a case and shed some insight into the case - specificity of a clinician's diagnostic performance found elsewhere (e.g., Elstein et. al., 1978). Secondly, they suggest a sensitivity in less experienced clinicians to specific training experiences, e.g., training materials, particular patient cases, etc. As experience increases, so does the sample of "inputs" and the effects of particular experience might be expected to lessen.

Case 2: Total Anomalous Pulmonary Venous Connection

The purpose of this case is to investigate the aggregation by subjects of a set of physiologically similar diseases into a memory grouping or category. The case is different from Case 1 in that while Case 1 dealt with a set of variants of one disease, Case 2 is concerned with a set of diseases.

The Logical Competitor Set for Case 2 includes four diseases: Total Anomalous Pulmonary Venous Connection (TAPVC), Partial Anomalous Pulmonary Venous Connection (PAPVC), Atrial Septal Defect (ASD), and Endocardial Cushion Defect (ECD).¹¹ Figure 6 shows the anatomical and physiologic abnormalities within the heart which define each of these diseases.

In TAPVC, all four pulmonary veins (PVn in Figure 6) connect to the right atrium (RA) of the heart rather than to the left atrium (LA), their normal site of connection. All oxygenated blood coming back to the heart from the lungs mixes with deoxygenated blood coming back to the heart from the "body". Hence,

¹¹

All of these diseases have variants as in Case 1. Variants of these diseases were treated as equivalent for Case 2 since interest in this case was in memory clustering at the disease level.

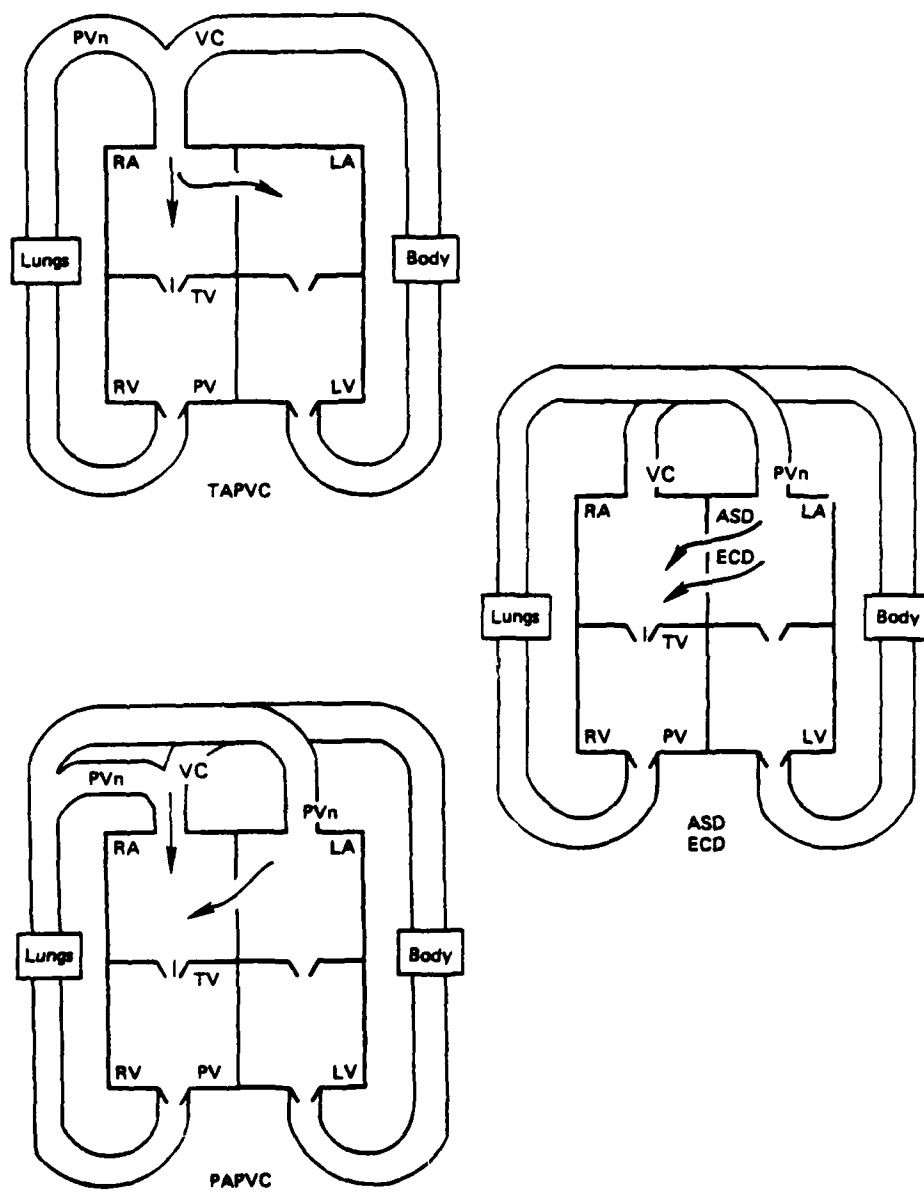


Figure 6. Logical competitor set for Case 2: TAPVC, PAPVC, ASD, ECD.

all blood subsequently pumped back to the body is a mixture of oxygenated and deoxygenated blood which causes the patient to appear cyanotic, that is, to take on a mildly "blue" skin coloration.

In PAPVC, only a subset of the pulmonary veins connect abnormally to the right atrium, with the remainder connecting, as they should, to the left atrium. A result is that some already oxygenated blood is recirculated through the lungs. Blood pumped to the body, however, is oxygenated and the patient retains his normal "pink" coloration.

Both ASD and ECD consist of a defect (a hole) in the atrial septum of the heart. They differ in the particular site of defect; ASD is a defect in the "upper" portion of the septum (the ostium secundum) while ECD is a defect in the lower portion of the septum (the ostium primum). In both diseases, the presence of the "hole" in the septum allows blood to shunt from the left atrium to the right atrium. While some oxygenated blood shunts to the right side to be recirculated to the lungs, blood expelled to the body is oxygenated and the patient is pink. Both TAPVC and PAPVC may also contain a hole in the atrial septum; this defect is necessary in TAPVC and optional in PAPVC.

A feature common to all four diseases in the LCS is an increased volume of blood in the right-sided chambers of the heart. This common element is a candidate feature on which diagnosticians might base a disease category, e.g., "diseases with right-sided volume overload". A clinical manifestation related to volume overload that all these diseases produce in common is a set of three auscultation findings. One is a murmur associated with increased blood flow across the tricuspid valve (TV). The second is a murmur associated with increased flow across the pulmonary valve (PV). The third is wide, fixed splitting of the second heart sound. The third finding is nearly pathognomonic for conditions of this type.

Of the four diseases, ASD is more common than the others. Hence, it might be expected that subjects' knowledge for this disease would develop more rapidly than for the others. More importantly, ASD is the disease that is used instructionally to introduce the concepts of atrial level left-to-right shunting of blood in the heart and right-sided volume overload. Therefore, it might be expected that the three auscultation findings (especially the "splitting") reflecting overload would be more strongly associated with ASD than with the other diseases. TAPVC, however is the

operative disease in the case.

There are six particularly important data items in the "patient file" presented to subjects for Case 2 (Appendix B, Case 2). Data items 17, 18, and 19 contain the set of three findings discussed above which are salient results of increased right-sided heart flow. Item 17 contains the "wide, fixed split second heart sound". Hence, it was expected that all subjects would raise at least ASD, the classic instance of this type of disease, by the time of these data points. Data item 7 (also 11), which reports that the patient is cyanotic, represents disconfirmatory evidence for all members of the LCS except TAPVC. Data item 21, which contains an x-ray description of "an unusual vascular shadow on the right side" is evidence against ASD and simultaneously constitutes a classic cue for PAPVC. In fact, one variant of PAPVC, "scimitar syndrome", derives its name from its presentation of such a finding on x-ray (Lucas & Schmidt, 1977, pp. 442). The EKG, item 22, contains a finding of "right axis deviation" on the EKG and constitutes strong disconfirmatory evidence for ECD. All data of the case are compatible with the operative disease, TAPVC.

Use of the Logical Competitor Set - Case 2

Table 4 shows all uses by all subjects of the four diseases in the Logical Competitor Set for Case 2 at all patient data points (Appendix B, Case 2) where any of the four was mentioned by any subject. The nature of this table is the same as Table 1 in that an "x" intersecting a disease and data item indicates that the subject mentioned the disease in his protocol at the time of the data item.

For reasons discussed above, it was assumed that most subjects would consider ASD in relation to the three data items, 17, 18, and 19. The use of other LCS members at these points is taken as evidence that the other diseases are associated in memory with ASD and this set of cues. Table 4 shows a decrease from students to experts, in the number of subjects who considered only ASD at these points. All of the students considered only ASD, the disease we presume to be the classic exemplar of right sided volume overload, at data items 17-19. Three of four trainees (T1, T2, T3) also considered only ASD while the fourth considered both ASD and TAPVC. The two least experienced experts also considered only ASD. Of the two highly experienced

Table 4
Case 2: Subject Use of LCS Hypotheses in Response to Patient Data Items

Subjects/Hypotheses		Patient Data Items ^a											
		History				Physical Exam				X-ray		EKG	
		1	3	5	7	HHx	17	18	19	20	HPEx	21 HXray	22 HEKG
S1	ASD					X	X				X		
	PAPVC											X	X
S2	ASD						X		X		X	X	
	ECD												X
	PAPVC											X	
S3	ASD						X				X		X
	ECD												X
	PAPVC											X	X
S4	ASD						X	X	X		X		X
	ECD												X
	TAPVC										X		
T1	ASD						X		X				
	PAPVC										X	X	X
T2	ASD										X		
	ECD								X		X		
	TAPVC											X	X
T3	ASD		X	X			X	X	X		X		X
	PAPVC												X
	TAPVC												X
T4	ASD						X			X			X
	ECD								X	X	X	X	X
	PAPVC												X
	TAPVC						X				X		X
E1	ASD	X	X	X		X	X		X	X			
	PAPVC										X	X	X
E2	ASD								X				X
	PAPVC										X		X
	TAPVC												X
E3*	ASD	X					X	X	X				
	ECD					X					X		X
	PAPVC						X	X					
	TAPVC					X	X	X	X		X	X	X
E4*	ASD						X		X		X		
	ECD								X				
	PAPVC										X		
	TAPVC										X		X

Note. (X) indicates a subject's use of a hypothesis at the time of a patient data item.

^a Statements of numbered patient data items are given in Appendix B, Case 2.

(HHx) etc. refers to points in the case where subjects are asked for hunches.

* The two experts with more than twenty years of experience.

experts, E3 utilized three LCS members (ASD, PAPVC, TAPVC) and E4 used two (ASD, ECD) at these points.

From the point of view of the entire case, no students, one trainee (T4), and two experts (E3, E4) generated all four members of the LCS during the course of the case. While this shows no obvious general trend toward increased use of the LCS with experience, it is perhaps significant that the full competitor set was used by the two high-level experts, E3 and E4.

In utilizing the full Logical Competitor Set, the two most experienced subjects, E3 and E4, demonstrated the same patterns of "precaution" and "extraction" respectively as they did in Case 1. E3 considered three of the four LCS members (ASD, PAPVC, TAPVC) at item 17, the first strong cue for right sided volume overload. E4 raised only ASD at this point and maintained this hypothesis until data item 21 which contains strong evidence against ASD. At this point, he expanded to the remainder of the LCS.

Diagnostic Errors - Case 2

Table 5 gives the final primary diagnoses for all subjects on Case 2. Only four subjects, trainees T2 and T3, and the two most experienced experts, E3 and E4,

Table 5
Case 2: Total Anomalous Pulmonary Venous Connection – Final Diagnoses

Subjects		Diagnosis
Students	S1	Partial Anomalous Pulmonary Venous Connection
	S2	Transposition of the Great Vessels + Pulmonary Stenosis + Atrial Septal Defect + Partial Anomalous Pulmonary Venous Connection
	S3	Endocardial Cushion Defect
	S4	Pulmonary Stenosis + Atrial Septal Defect + Ventricular Septal Defect
Trainees	T1	Partial Anomalous Pulmonary Venous Connection
	T2	Total Anomalous Pulmonary Venous Connection
	T3	Total Anomalous Pulmonary Venous Connection
	T4	Atrial Septal Defect
Experts	E1	Partial Anomalous Pulmonary Venous Connection
	E2	Partial Anomalous Pulmonary Venous Connection
	E3*	Total Anomalous Pulmonary Venous Connection
	E4*	Total Anomalous Pulmonary Venous Connection

* The two experts with more than twenty years of experience.

diagnosed the case correctly. Two of these four subjects are ones who considered the full LCS in the case. Subjects who diagnosed the case incorrectly demonstrate informative types of errors.

Student S3 diagnosed the case as Endocardial Cushion Defect (ECD). The strongest evidence against this disease is the finding of "right axis deviation" on the EKG (item 22). ECD uniformly presents with left axis deviation and, in fact, is one of a very few congenital heart diseases that does; hence, left axis deviation is a nearly pathognomonic finding for ECD. S3 not only evaluated the "right axis" as positive evidence for ECD, but, in addition "triggered" or proposed ECD for the first time at this point (see Table 4). This is, simply, imprecision in the subject's disease model for ECD. It is as though the subject remembered that the EKG axis is important in ECD but could not remember the details.

The final diagnosis of subject T4 was ASD, even though she had considered TAPVC during the case. She correctly evaluated cyanosis (blueness - items 7 and 11) as negative for ASD, but maintained ASD nonetheless. Her primary difficulty was that she did not believe

that TAPVC could present in a child as old as the one in the case (5 years old) although it certainly can - as the case itself, a real case, attests. This suggests that the allowable age range specified in the subject's disease model for TAPVC is overly restrictive, probably reflecting a limited sample of experiences with this disease.

Four subjects (S1, T1, E1, E2) diagnosed the case as PAPVC. Three of these subjects (S1, T1, E1) show a pattern in which only ASD (among the LCS members) is considered prior to item 21, a classic x-ray cue for PAPVC, and only PAPVC is considered at that point and thereafter (see Table 4). This indicates a strong data-driven dependence in the diagnosis by these subjects; that is, the subjects are pushed from hypothesis to hypothesis depending on the most recent strong disease cue in the data, and when new hypotheses are generated, these are not strongly enough associated in memory with other LCS members to activate these other diseases. Some support for this claim can be seen in subject T1's protocol taken from the point in the case where he offers his final diagnosis:

T1: I am sort of drawing a blank on how to fit all this information together. And ah, I am just sort of guessing right now. I would say just Scimitar Syndrome (PAPVC) primarily based on the chest x-ray and ah, I'm not really sure whether the whole thing fits together well. That is all I can say.

Of the four subjects, student S1 never evaluated PAPVC with respect to cyanosis; hence, this finding had no opportunity to detract from his PAPVC hypothesis. Subject T1 evaluated cyanosis as confirmatory evidence for PAPVC and this erroneous evaluation reinforced this disease interpretation. Expert subjects E1 and E2 evaluated cyanosis appropriately as negative evidence for PAPVC, but this evaluation was probably overridden by the strength of the cue for PAPVC on the x-ray.

Finally, two students (S2, S4) proposed configurations of multiple diseases as explanation for the case. Both of these composite diagnoses included the disease Pulmonary Stenosis (PS) and it is this component of the final diagnosis that is the key to understanding the performance of these two subjects. Table 6 shows the interpretations by all subjects of data item 18, a "systolic murmur" in auscultation of the heart. Such a murmur results whenever there is too much flow

Table 6
Case 2: Interpretations of Data Item 18

Subject		Interpretation	
		Pulmonary Stenosis	Increased Flow Pulmonary Valve
Students	S1	+	
	S2	+	
	S3	+	+
	S4	+	
Trainees	T1		
	T2	+	
	T3	+	+
	T4		+
Experts	E1	+	+
	E2		
	E3*		+
	E4*	+	+

Note. (+) indicates that a subject interpreted the murmur of data item 18 as Pulmonary Stenosis or increased flow over the pulmonary valve.

* The two experts with more than twenty years of experience.

over the pulmonary valve, relative to its orifice size. This situation prevails in either of two conditions: (1) When there is normal amount of flow but an abnormally small orifice. This is the disease Pulmonary Stenosis which refers to an abnormally tight valve. (2) When there is a normal sized orifice but abnormally high flow, the situation which prevails in the diseases of the LCS. A "+" under one of these two interpretations in Table 6 indicates that a subject attributed this interpretation to the murmur of data item 18.

Table 6 shows that most of the students (3 of 4) interpreted the murmur only as Pulmonary Stenosis while most of the expert group (3 of 4) interpreted the murmur as increased flow or a tight valve. While student S1 (also subject T2) was eventually able to extract himself from his interpretation, students S2 and S4 were not. Once these students introduced PS into their diagnoses, they were forced to propose rather unusual combinations of multiple diseases to account for some of the findings of the case. For example, subject S2, in order to reconcile PS with other data of the case indicating increased blood flow in the lungs, simply transposed the great vessels of the heart; that is, he

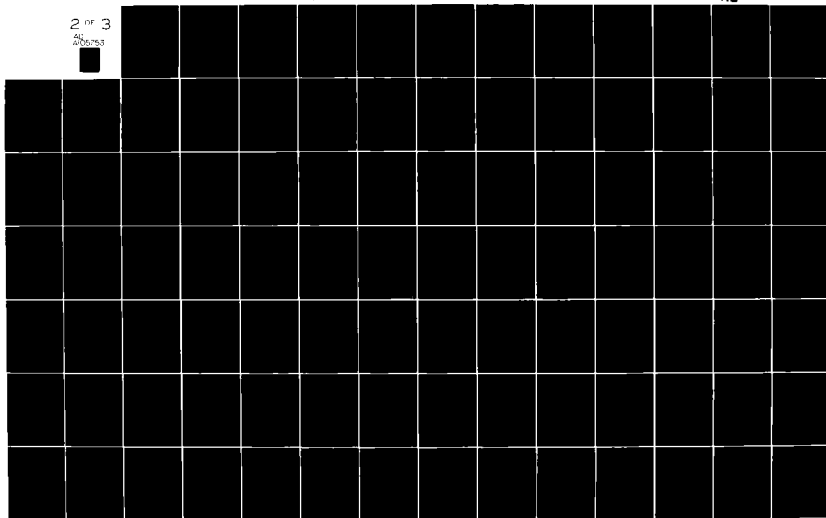
detached the pulmonary artery from its normal mooring at the pulmonary valve, and reattached it at the aortic valve and did the opposite with the other great vessel, the aorta. While this transposition is a congenital heart disease, Transposition of the Great Vessels, it is highly unlikely that a child with the combination of abnormalities proposed by the subject could have lived for five years untreated.

The interpretations of the systolic murmur by the students in Case 2 is another example of error, or at least imitation, in causal knowledge. It represents a situation where there are multiple causes for a finding and the novice considers only a subset. This is not unlike what has been shown at the disease and disease variant levels; that is, when multiple diseases in the Logical Competitor Set can produce a finding, the novice seems limited to the most salient members. This suggests the import of grouped or clustered memory organization not only for diseases, but also for "low-level, pathophysiologic interpretations for data.

AD-A105 753 PITTSBURGH UNIV PA LEARNING RESEARCH AND DEVELOPMEN--ETC F/G 6/5
KNOWLEDGE BASED COMPONENTS OF EXPERTISE IN MEDICAL DIAGNOSIS.(U)
UNCLASSIFIED SEP 81 P J FELTOVICH N00014-79-C-0215
LRDC-81/PDS-2 NL

2 OF 3

AL
000003



Case 3: Patent Ductus Arteriosus

The purpose of this case is to test the robustness of expert grouping of hypotheses in a straightforward case in which there are no data discrepant with an initially induced disease interpretation. Interest is in whether subjects, even in a case with a highly common disease, strong cues for this disease, and no data discrepant with this interpretation, still investigate a related set of physiologically similar alternatives.

The operative disease in the case is Patent Ductus Arteriosus (PDA), a schematic for which is shown in Figure 7. This disease is an extra-cardiac shunt, that is, an abnormal communication between vessels, the aorta (Ao) and pulmonary artery (PA), outside the heart. There are four other "disease" conditions in the Logical Competitor Set. Congenital heart diseases, Arterio-venous Fistula (AVF) and Aorto-Pulmonary Window (APW) are other extra-cardiac shunts. Venous Hum (VH) is a benign condition that presents a murmur similar to PDA, and Ruptured Sinus of Valsalva (RSV) is a heart condition that has a similar clinical presentation to PDA.

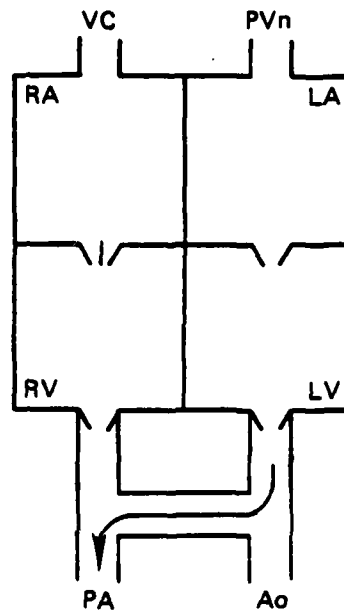


Figure 7. Patent Ductus Arteriosus.

In the patient file presented to subjects for Case 3 (Appendix B, Case 3), the most important patient data item is number 19, a classic murmur of Patent Ductus Arteriosus. It was assumed that all subjects would generate PDA as a hypothesis no later than this point. No data of the case are incompatible with PDA.

Use of the Logical Competitor Set - Case 3

Table 7 shows all uses of members of the Logical Competitor Set by all subjects during the course of the case. (It is like Table 1; only the diseases recorded and the referents for the "patient file" data items listed across the top are changed). It is clear that only one subject, E3, one of the two high-level experts, considered the full competitor set, although expert E2 considered three of the five - more than any of the remaining subjects.

If one inspects the subjects who used even one additional LCS member other than PDA, it is found that three students (S1, S3, S4), one trainee, T1, and three experts (E1, E2, E3) meet this criterion. Most student uses of a second LCS member were isolates, so that if one requires that the additional member be used more than once, this condition holds for only one student, S3, one trainee, T1, but still three of the

Table 7
Case 3: Subject Use of LCS Hypotheses in Response to Patient Data Items

Subjects/Hypotheses		Patient Data Items ^a												
		History					Physical Exam				X-ray		EKG	
		3	4	5	7	HHx	14	19	20	HPEx	21	HXray	22	HEKG
S1	PDA					X		X	X	X	X	X	X	X
	APW												X	
S2	PDA					X		X	X	X	X	X	X	X
S3	PDA		X	X	X	X		X	X	X	X	X	X	X
	AVF								X					X
S4	PDA					X	X	X	X	X	X	X	X	X
	AVF						X							
T1	PDA						X	X	X	X	X	X	X	X
	APW						X	X	X		X			X
T2	PDA							X	X	X	X	X	X	X
T3	PDA							X	X	X	X	X	X	X
T4	PDA	X					X	X	X	X	X	X	X	X
E1	PDA						X	X	X	X	X			X
	AVF							X			X			X
E2	PDA						X	X	X	X	X	X	X	X
	AVF						X	X						
	VH							X	X					
E3 ^o	PDA							X	X	X			X	
	AVF								X					
	VH							X						
	APW							X	X	X				
	RSV								X					
E4 ^o	PDA							X	X		X	X	X	

Note. (X) indicates a subject's use of a hypothesis at the time of a patient data item.
^a Statements of numbered patient data items are given in Appendix B, Case 3.
 (HHx) etc. refer to points in the case where subjects are asked for hunches.
 *The two experts with more than twenty years of experience.

experts (E1, E2, E3).

Expert E3 considered the full LCS in a "pre-cautionary" pattern consistent with his performance on other cases (see Table 7). He used three of the five LCS members as hypotheses at data item 19, a strong cue for PDA. The remainder of the LCS was filled out two items later, after an intervening, uninformative data item, at the point where the subject was asked for "hunches". The other high-level expert, E4, looks in all respects like a novice in this case in that he considered only PDA. However, if our earlier interpretations of an "extraction" method are correct for this subject, we would not expect him to expand to other members of the competitor set unless he encountered data discrepant with his target hypothesis; of course, there are none in this case.

The diseases in this case constitute a category of "Extra-Cardiac Communications" and related conditions. An interpretation of the results of this case is that with high-level experience, it is this category, and not isolated individual members, that is generated and tested when a strong cue for a category member is encountered. No subject diagnosed this case incorrectly, hence analysis of subject errors is uninformative.

Case 4: Pulmonary Atresia

The objective of this case is similar to that of Case 2; that is, to assess subjects' aggregation of physiologically similar diseases into categories. Case 4 is different from Case 2 in that, unlike Case 2, no single cue serves to distinguish the members of the Logical Competitor Set from diseases outside it (as did "wide, fixed split second heart sound" in Case 2). In Case 4 the diagnostician must arrive at the LCS by partitioning the space of diseases, using multiple data items from widely separated parts of the case.

The Logical Competitor Set for Case 4 includes three diseases, Pulmonary Atresia (PAT), Tricuspid Atresia (TAT), and Ebstein's Malformation of the Tricuspid Valve (EBST).¹² Figure 8 depicts the anatomical abnormalities within the heart which define each of these diseases. In Pulmonary Atresia and Tricuspid Atresia, the pulmonary and tricuspid valves respectively are "shut" (only tissue exists where the valves should be). In Ebstein's disease, a diminutive and

¹²

The abbreviations PAT and TAT will be used instead of the more conventional abbreviations PA and TA in order to avoid confusion between "Pulmonary Atresia" and "pulmonary artery" which we have abbreviated PA.

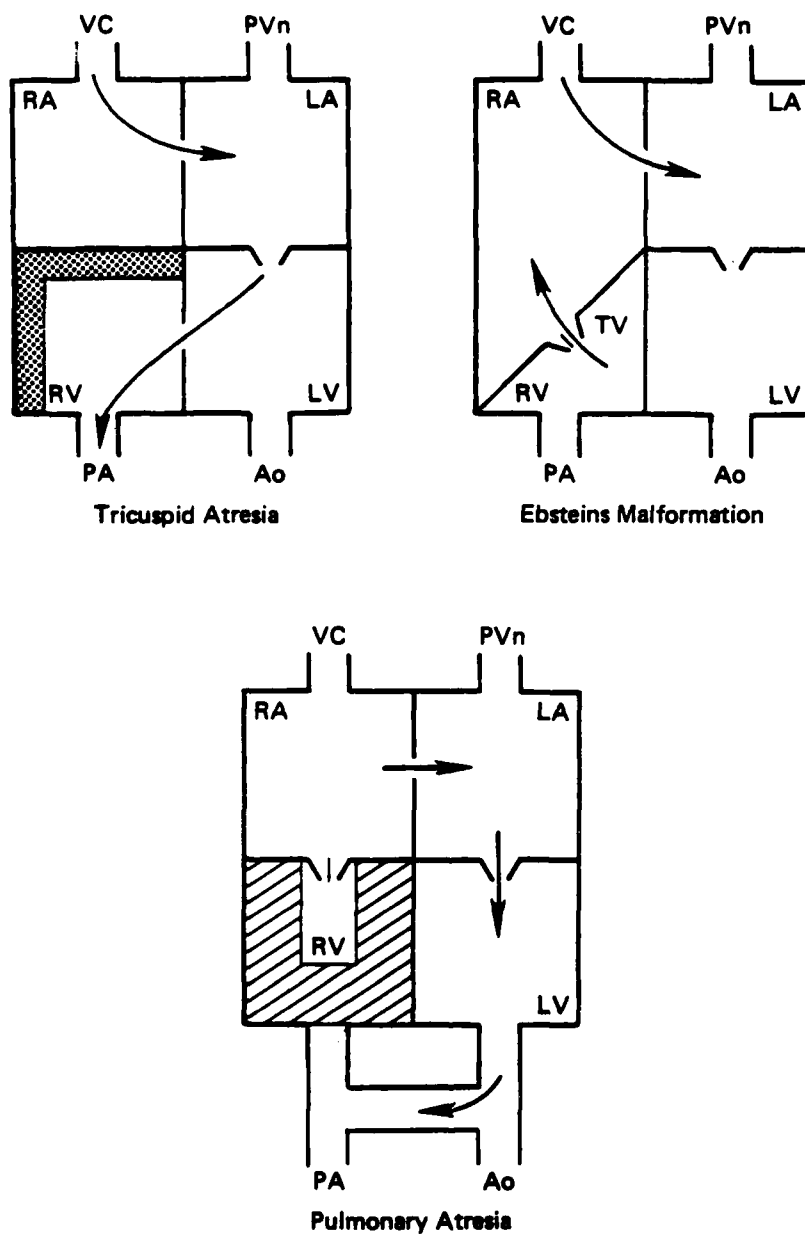


Figure 8. Logical competitor set for Case 4: Pulmonary Atresia, Tricuspid Atresia, Ebstein's Malformation.

non-compliant right ventricle (RV) restricts inflow of blood to that ventricle and an incompetent tricuspid valve (TV) allows some regurgitation of blood from the right ventricle back to the right atrium(RA). The net physiology of all these diseases is one of obstruction to blood flow on the right side of the heart resulting in reduced blood flow to the lungs and right-to-left shunting of blood at the atrial level within the heart. The right-to-left shunting and diminished blood flow to the lungs cause the patient to be cyanotic (blue skin coloration). In short, these diseases constitute a physiologic category of "cyanotic diseases with decreased pulmonary blood flow."

Pulmonary Atresia is the operative (or true) disease in the case. The three members of the LCS are best discriminated on the EKG. Tricuspid Atresia produces a finding of "left axis deviation" on the EKG while Pulmonary Atresia produces a normal EKG axis. Ebstein's, unlike the other two, produces an EKG finding of "right bundle branch blocking." All other clinical manifestations of the three diseases are quite similar.

There are several key data items in the "patient file" presented to subjects for Case 4 (Appendix B, Case 4). The subject receives evidence of cyanosis during history and early physical examination (1, 3, 8). The x-ray, item 17, contains evidence of diminished blood flow to the lungs and, with the cyanosis evidence, could enable the subject to narrow diagnosis to the three members of the LCS. The EKG, item 18, contains information to discriminate among these.

Use of the Logical Competitor Set - Case 4

Table 8 shows all uses of members of the Logical Competitor Set by all subjects during the course of the case. (It is like Table 1; only the diseases recorded and the referents for the "patient file" data items listed across the top are changed.)

Table 8 shows a clear increase in the use of the full Logical Competitor Set from students to trainees, but no clear difference in this regard between trainees and experts. In particular, no student considered the full LCS, and two students (S1, S3) considered only one member. All four trainees and three experts (E1, E2, E3) used all of the diseases in the LCS. Two experts (E2, E3) used all three diseases more than once while no trainee did - suggesting

Table 8
Case 4: Subject Use of LCS Hypotheses in Response to Patient Data Items

Subjects/Hypotheses		Patient Data Items ^a							
		History	Physical Exam				X-ray		EKG
		HHx	14	15	16	HPEx	17	Hxray	18 HEKG
S1	TAT						X	X	X
S2	TAT			X				X	X X
	PAT			X	X	X		X	X
S3	EBST								X X
S4	TAT						X	X	X X
	PAT			X		X	X	X	X
T1	EBST						X		
	TAT			X		X	X	X	X X
	PAT					X			
T2	EBST								X
	TAT								X
	PAT			X					
T3	EBST						X		
	TAT			X					X
	PAT			X	X	X	X	X	X X
T4	EBST								X
	TAT						X		X
	PAT						X		X X
E1	EBST						X		
	TAT						X	X	X X
	PAT	X		X		X	X	X	X X
E2	EBST						X	X	
	TAT						X	X	X X
	PAT						X	X	X
E3*	EBST						X	X	X X
	TAT		X				X	X	X X
	PAT							X	X X
E4*	PAT						X	X	X X

Note. (x) indicates a subject's use of a hypothesis at the time of a patient data item.

^a Statements of numbered patient data items are given in Appendix B, Case 4.

(HHx) etc. refer to points in the case where subjects are asked for hunches.

* The two experts with more than twenty years of experience.

somewhat more active consideration of the Logical Competitor Set by these experts.

While both trainees and experts considered the full Logical Competitor Set, their patterns of use of these diseases were different. Three of the four experts used all members of the LCS at data point 17 (the x-ray) or at the immediately succeeding point where subjects reported hunches. Since item 17 is the data item that allows specification of the category "cyanotic heart disease" into the category "cyanotic diseases with decreased pulmonary blood flow," this pattern suggests that the expert subjects were using this category. In contrast, no trainees used all three LCS members at either of these points, suggesting that these three diseases do not, at least to the same extent, constitute a functional diagnostic category for these subjects.

Regarding the expert diagnostic modes of "precaution" and "extraction," expert E3 again considered all three LCS members together before the onset of data useful for discriminating among them. Expert E4 considered explicitly only Pulmonary Atresia, the correct disease, at data item 17. However, his protocol from the immediately succeeding data point, HXray (hunches

after x-ray), shows explicit consideration of the category of "cyanotic disease with decreased pulmonary blood flow" with targeting for active consideration of the particular LCS member he judged most likely:

Experimenter: At the end of x-ray can you tell me about hunches?

E4: At this point the picture would be more likely that of cyanotic heart disease involving decreased pulmonary blood flow. The specific defect would seem to be pulmonary atresia with intact septum.

Since no succeeding data are discrepant with his target hypothesis, his performance is consistent with the "extraction" mode as we have proposed it. In addition, E4's overt consideration of the LCS category here lends credence to a speculation we have made about the "extraction" mode in Case 1 and Case 2, that is, that the subject covertly considered the LCS category in those cases before he overtly articulated the members.

Diagnostic Errors - Case 4

Table 9 gives the final primary diagnoses for all subjects. The final diagnoses of the students on this case are outside the LCS and the full explanation for their performance is not transparent. However, a

Table 9
Case 4: Pulmonary Atresia – Final Diagnoses

Subjects		Diagnosis
Students	S1	Truncus Arteriosus
	S2	Hypoplastic Right Ventricle
	S3	Truncus Arteriosus
	S4	Pulmonary Atresia
Trainees	T1	Tricuspid Atresia
	T2	Tricuspid Atresia
	T3	Tricuspid Atresia
	T4	Pulmonary Atresia
Experts	E1	Pulmonary Atresia
	E2	Pulmonary Atresia
	E3*	Ebstein's Malformation
	E4*	Pulmonary Atresia

*The two experts with more than twenty years of experience.

partial explanation can be given.

Two students (S1, S3) gave a final diagnosis of Truncus Arteriosus. Truncus is a congenital heart disease in which the aorta and pulmonary artery, the two great vessels which normally lead out of the heart, are merged into one large outlet vessel. This single vessel is associated with only one outlet valve from the heart. The single valve results in a patient finding of "single second heart sound" on auscultation since these "heart sounds" are produced by the closing of heart valves; in a normal heart there are two components to the second heart sound.

While Truncus produces a single heart sound, so do a number of other diseases, including all members of the Logical Competitor Set. It is not even necessary that only one valve exist for only a "single sound" to be produced; the same finding is produced when there are two outlet valves but the blood flow across one of them is substantially diminished - the situation in Ebstein's and Tricuspid Atresia.

One explanation for the performance of students S1 and S3 is that they judged the "single sound" to be more discriminating for Truncus than it really is. Some evidence for this interpretation can be seen in

protocols from these two subjects showing interpretations of the patient finding of a "single sound" (Figure 9). It is clear that this finding had a substantial influence on the final diagnoses of these subjects.

If our interpretation for these subjects is true, it would be another example of how the beginning practitioner is restricted in the number of alternative explanations he can bring to bear on a finding, either at the level of alternative pathophysiological causes or alternative disease explanations. Whether "single second sound" is a more "classic" cue for Truncus than for other diseases is open to question. However, it would seem reasonable that the "common trunk" which defines this disease would serve to accentuate the single sound as an expected finding in the disease.

S2, the other student who misdiagnosed Case 4, gave as a final diagnosis (Hypoplastic Right Ventricle) one of the patient findings presented in the case (the EKG); that is, the subject used one of the patient data items as a final diagnosis. An explanation for this student's performance will not be ventured.

The trainees and experts are nicely split on this case, with most trainees (3 of 4) judging Tricuspid Atresia and most experts (3 of 4) judging

(15) The second heart sound is single and perhaps slightly increased in intensity. There is no gallop or diastolic murmur.

S1: Well, this is a significant finding because ah, the fact that the second heart sound is not split ah, suggests that ah, we would be dealing with a truncus.

S3: It could be ah, ah. There is a single outflow tract ah. It could be truncus arteriosus. Ah, that would fit with the single S2 (second heart sound) . . . So, I'll go with number one on my list as ah, truncus arteriosus and I'm not sure what type. I'd have to do an angio. I guess, or I mean arteriography.

Figure 9. Protocols from subjects S1 and S3 showing interpretations of "single second heart sound" — Case 4.

Pulmonary Atresia, the correct disease. Recall that TAT and PAT are distinguishable on the axis of the EKG where TAT presents "left axis deviation" and PAT presents a normal, undeviated axis. It is on the subjects' evaluations of this particular data item that we might expect to find an explanation for the performance of these two groups.

Table 10 shows all explicit evaluations by subjects of the EKG axis as confirmatory (+), disconfirmatory(-), or ambivalent(o) evidence with respect to Pulmonary Atresia and Tricuspid Atresia. All subjects below the expert level, who explicitly evaluated the axis with respect to either of these two diseases, evaluated the axis as confirmatory evidence for Tricuspid Atresia. Subject T1 proceeded to conclude TAT as a final diagnosis. While subjects S4 and T4 ultimately chose Pulmonary Atresia over Tricuspid Atresia on other grounds, their explicit evaluations of the axis as confirmatory for TAT contribute to an explanation of what happened to other non-expert subjects who maintained Tricuspid Atresia. In particular, even though T2 did not overtly evaluate the EKG axis with respect to Tricuspid Atresia, immediately after the EKG she generated TAT for the first time in

Table 10
Case 4: Evaluations of EKG Axis in Relation to Tricuspid Atresia
and Pulmonary Atresia

Subjects		Hypotheses	
		Tricuspid Atresia	Pulmonary Atresia
Students	S1		
	S2		
	S3		
	S4	+	
Trainees	T1	+	
	T2		
	T3		
	T4	+	
Experts	E1	0	+
	E2	-	
	E3*	-	
	E4*		

Note. (+, -, 0) indicate that the subject evaluated the EKG axis as confirmatory, disconfirmatory, or ambivalent evidence respectively in relation to the hypothesis.

* The two experts with more than twenty years of experience.

the case as her only "hunch". Likewise, subject T3 raised TAT at this same point as a "hunch" after not having used this hypothesis since early physical examination (see Table 8). All expert subjects who explicitly evaluated the axis, evaluated it as either disconfirmatory for Tricuspid Atresia or confirmatory for Pulmonary Atresia.

The EKG axis as presented in the case is +50 degrees which technically represents left axis deviation (for a 4 day old child, as presented in the case - Moller 1978, p. 24) as one would expect in Tricuspid Atresia, so that if one were using the "text-book" rule for discriminating PAT and TAT (e.g., Moller, 1978, p. 137), Tricuspid Atresia would be the diagnosis of choice in the case. However, the expert evaluations of this finding, as well as post-experimental discussions with these subjects, confirmed that the experts judged +50 degrees to be "just not far enough leftward" for Tricuspid Atresia and that these subjects would require the axis to be "down around zero or negative" before they would choose TAT over PAT. We see here a nice example of overly general, textbook-like rules of evaluation and clinical expectations in

less experienced subjects (imprecise disease models) and pin-point refinement of these in more experienced diagnosticians, probably just reflecting their greater clinical experience with the two diseases and their manifestations.

Case 5: Coarctation of the Aorta
Mitral Insufficiency, Bicuspid Aortic Valve

This case was chosen as a test of subjects' ability to detect the need for and to include in a diagnosis additional abnormalities beyond a strongly induced initial disease interpretation. Unlike other cases in the study, logical competitors for the principle diagnosis are not in question; the main disease interpretation is clear-cut. What is at issue is the addition of other abnormalities to this interpretation.

The case is one in which there is a very strong early cue for the principle diagnosis, Coarctation of the Aorta (Coarc); this disease has a pathognomonic cue which appears among the patient findings presented in Case 5. In our initial conception, later findings were believed to be inconsistent with this interpretation alone, requiring not a switch to another disease nor to a variant of Coarctation, but rather the conjoining of two additional diseases to the original Coarctation interpretation. The two additional abnormalities believed necessary were Mitral Insufficiency (MI) and Bicuspid Aortic Valve (BAV, a kind of Aortic Stenosis). The case was seen as a test of the precision of subjects' diseasemodels for the initially induced disease, Coarctation; subjects would need to recognize the

discrepancy between this disease, taken alone, and certain findings of the case. As will be seen, the results from this case are sufficiently complicated, perhaps counter intuitive, that a discussion of the background of our prior analysis seems warranted.

For the other four cases included in the study the "correct" answer, that is, the correct final diagnosis was established by cardiac catheterization, an invasive, probing procedure of the heart which is generally definitive. For the present case, no catheterization was ever carried out. For this reason, the correct diagnosis for this case was not available in the same sense as for the other cases of the study. Alternatively, we relied on the following body of external evidence for establishing our analysis of the case:

- 1) When the patient represented in the case was originally diagnosed at the University of Minnesota Hospitals, the three-part diagnosis of Coarctation, Mitral Insufficiency, and Bicuspid Aortic Valve was made.

- 2) The patient was later rediagnosed at the Mayo Hospitals in Rochester, Minnesota where the same diagnosis was made.

- 3) On two different occasions, separated by a year, the expert consultant to the project diagnosed the case

in the same experimental format and with the same patient data used in the patient study. On both occasions, his final diagnosis included all of the components, that is, Coarc, MI, and BAV. While our interpretation of the case is not based on a catheterization, we continue to believe that it is sound.

Figure 10 depicts the anatomic abnormalities within the heart that define the correct diagnosis for Case 5. Coarctation of the Aorta is a constriction in the aorta (Ao), down-stream of the left ventricle (LV). This constriction has a "stenotic" effect on blood flow; that is, it increases resistance to blood flow and forces the left ventricle to work harder to eject blood. Bicuspid Aortic Valve is a condition in which the aortic valve (AV) has two cusps rather than three. While there can be a subtle semantic and physiologic distinction between BAV and valvular "Aortic Stenosis", the two conditions are so similar that they were treated as equivalent in all analyses of the case. Mitral Insufficiency is a condition in which the mitral valve (MV), between the left atrium (LA), and left ventricle, remains partly open during the phase of the heart cycle when the left ventricle is ejecting blood. During this phase a normal mitral valve is closed, preventing back-flow of left ventricular blood to the left

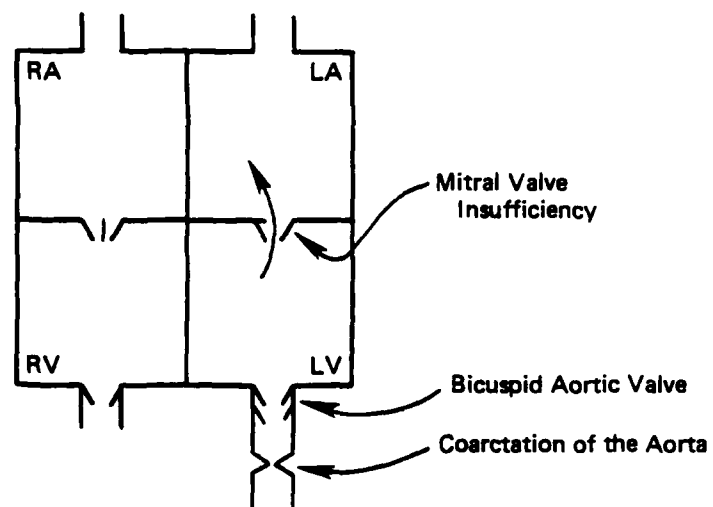


Figure 10. Components of the correct diagnosis for Case 5: Coarctation of the Aorta; Bicuspid Aortic Valve, Mitral Insufficiency.

atrium. An "insufficient" mitral valve allows regurgitation or back-flow of blood into the left atrium. This regurgitant blood causes the left atrium to enlarge and also produces a murmur.

In the "patient file" presented to subjects for Case 5 (Appendix B, Case 5), data item 12 contains a description of substantial discrepancy between blood pressures taken in the arms and legs. This is very nearly pathognomonic for Coarctation and it was assumed all subjects would establish Coarctation as a hypothesis by this point in the case. Data items 18 and 20 are cues for Bicuspid Aortic Valve (Aortic Stenosis). Eighteen is a systolic ejection click. Twenty is a murmur of Valvular Aortic Stenosis. Data item 19 contains a rather classic description of the murmur of Mitral Insufficiency although the murmur could reasonably be interpreted as emanating from the Coarctation of the Aorta. Data items 22 and 23, the x-ray and the EKG respectively, contain three findings that, in our view, require that Mitral Insufficiency be included in an optimal diagnosis. These are findings of prominent pulmonary vascularity on x-ray, right ventricular hypertrophy on EKG and left atrial enlargement on both x-ray and EKG.

Use of the Components of the Correct Diagnosis - Case 5

Table 11 shows all uses by subjects of the three components (Coarc, MI, BAV) of the correct diagnosis during the course of the case. (The table is like Table 1; only the diseases recorded and the referents for the patient data items listed across the top are changed). An "x" in the intersection of a data item and disease component in this table indicates that the subject mentioned the disease as a hypothesis at that patient data point.

Table 11 shows that two students (S3, S4), all four trainees, and two experts (E2, E3) ever considered all three components of the correct diagnosis. From the point of view of subjects entertaining all of the correct components, regardless of whether or not they ultimately included these in a final diagnosis, it is clear that such simple consideration did not increase uniformly with experience.

Since all subjects generated Coarctation as a hypothesis, it is useful to examine separately the other two components of the correct diagnosis. Two students (S3, S4), all four of the trainees, and all of the experts used Bicuspid Aortic Valve as an additional hypothesis to Coarctation during the course

Case 5: Subject Use of the Components of the Correct Diagnosis as Hypotheses in Response to Patient Data Items

Table 11

Subjects/Hypotheses	Patient Data Items ^a																				EKG	
	History				Physical Exam												X-ray		EKG			
	1	4	HHx		12	13	14	15	16	17	18	19	20	HPE _x	22	HXray	23	HEKG				
S1 COARC MI					X			X	X	X	X		X	X		X	X	X				
S2 COARC MI					X	X			X	X	X		X	X		X	X	X				
S3 COARC BAV					X	X								X		X		X				
S4 COARC BAV MI					X	X	X				X	X	X	X		X	X	X				
T1 COARC BAV					X	X		X		X	X		X	X		X	X	X				
T2 COARC BAV											X		X	X		X	X	X				
T3 COARC BAV MI					X		X			X	X	X	X	X		X	X	X				
T4 COARC BAV MI					X		X						X	X		X	X	X				
E1 COARC BAV					X	X				X	X		X	X		X	X	X				
E2 COARC BAV					X	X	X			X	X	X	X	X		X	X	X				
E3*COARC BAV MI					X	X					X	X	X	X		X	X	X				
E4*COARC BAV					X	X	X		X		X	X	X	X		X	X	X				

Note. (X) indicates a subject's use of a hypothesis at the time of a patient data item. (HHx) etc. refer to points in the case where subjects are asked for hunches.
^a Statements of numbered patient data items are given in Appendix B, Case 5.

of the case. Clearly it is not the BAV component of the correct diagnosis that accounts for the somewhat counter-intuitive results for the case. The only subjects who failed to utilize the cues for BAV were students. The explanation for the unexpected results in the case is to be found in subjects' handling of Mitral Insufficiency. In particular, all four students, all trainees, but only two experts (E2, E3), ever generated this component as a hypothesis for the case. Moreover, expert E3 mentioned Mitral Insufficiency only once and can hardly be judged to have given this hypothesis serious contemplation.

Diagnostic Errors - Case 5

The experts in this study did not judge it necessary to add a component of Mitral Insufficiency to their Coarctation diagnosis although some less experienced subjects did. This is illustrated in the final diagnoses subjects gave for the Case (Table 12). Only three subjects (S3, T1, T4) offered the correct final diagnosis and Mitral Insufficiency was obviously the impediment, since no expert included it in the final diagnosis - although all experts included Bicuspid Aortic Valve.

Table 12
Case 5: Coarctation, Mitral Insufficiency, Bicuspid Aortic Valve – Final Diagnoses

Subjects		Diagnosis
Students	S1	Coarctation + Mitral Insufficiency
	S2	Coarctation + Ventricular Septal Defect
	S3	Coarctation + Mitral Insufficiency + Aortic Stenosis
	S4	Coarctation + Mitral Insufficiency
Trainees	T1	Coarctation + Mitral Insufficiency + Aortic Stenosis
	T2	Coarctation + Ventricular Septal Defect
	T3	Coarctation + Mitral Insufficiency + Bicuspid Aortic Valve + Atrial Septal Defect
	T4	Coarctation + Mitral Insufficiency + Bicuspid Aortic Valve
Experts	E1	Coarctation + Bicuspid Aortic Valve
	E2	Coarctation + Bicuspid Aortic Valve
	E3*	Coarctation + Bicuspid Aortic Valve
	E4*	Coarctation + Bicuspid Aortic Valve

* The two experts with more than twenty years of experience.

Although the case did not turn out as expected, it does offer the clearest discrimination between the expert group and other subjects of any case in the study. Ignoring the BAV component of the diagnosis, experts basically arrived at a "Coarctation only" diagnosis (BAV occurs in about 50% of Coarcs and might be thought of as a defining component of one variation of Coarctation). Since the case was originally designed as a test of augmentation to Coarctation in a diagnosis, the overall results beg an explanation.

Perhaps the best explanation for the general result lies in the interpretation of the murmur, data item 19, which was originally thought to be the strongest evidence for Mitral Insufficiency. Table 13 shows the interpretation of this murmur by subjects in relation to Mitral Insufficiency and Coarctation of the Aorta. A "plus" in this table indicates that a subject interpreted data item 19 as confirmatory evidence for the interpretation (COARC or MI) under which it is listed. A "zero" indicates an ambivalent interpretation of the murmur with regard to the column heading; an example of such an interpretation is given below:

T4: It doesn't say if this murmur is audible over the back which would be, ah, well occur, if this were just . . . if this were his murmur of the coarc.

Table 13
Case 5: Evaluation of a Murmur as Coarctation
of the Aorta or Mitral Insufficiency

Subjects		Interpretation of Murmur	
		Mitral Insufficiency	Coarctation of the Aorta
Students	S1	+	
	S2	+	0
	S3	+	
	S4	+	
Trainees	T1	+	
	T2		
	T3	+	0
	T4	+	0
Experts	E1		+
	E2		+
	E3*	+	+
	E4*		+

Note. (+) indicates that the subject judged the murmur description to be consistent with an interpretation of Mitral Insufficiency or Coarctation of the Aorta.

(0) indicates that the subject judged the murmur description to be ambivalent with regard to an interpretation.

* The two experts with more than twenty years of experience.

Seven of the eight subjects below the expert level interpreted the murmur given in item 19 as Mitral Insufficiency. Three of these raised the possibility that the murmur represented Coarctation, but this interpretation was clearly not their preferred choice. In contrast, all four experts interpreted the murmur solidly as emanating from the Coarctation of the Aorta. Expert E3 was the only expert who even raised the possibility of Mitral Insufficiency for the murmur. After this murmur, however, he never mentioned Mitral Insufficiency again. Although expert E2 eventually brought Mitral Insufficiency into his diagnostic thinking, his only explicit evaluation of the murmur of data item 19 was as Coarctation.

The evaluations of this murmur go a long way toward explaining the results of Case 5. Students and trainees generally had no problem adding Mitral Insufficiency into their case interpretation while experts did. By way of accounting for this result, Figure 11 shows three murmur descriptions. One of these is data item 19, the murmur description given in the case. The other two are introductory textbook (Moller, 1978) descriptions of the murmurs of Mitral Insufficiency and Coarctation of the Aorta. It seems that the murmur as presented in Case 5 is a "textbook" murmur of

(a) *Murmur – Data item 19 - Case 5*

There is a grade 3/6 soft systolic murmur at the apex which radiates to the axilla.

(b) *Textbook Mitral Insufficiency Murmur*

"An apical (apex) pan systolic murmur . . . This murmur radiates to the axilla and may be associated with a thrill."

(c) *Textbook Coarctation Murmur*

"An (systolic) ejection type murmur is present along the sternal border, at the apex, and over the back between the scapulae. The murmur is generally of grade 2-3/6 intensity."

Figure 11. Murmur as described in data item 19, Case 5(a). Textbook (Moller, 1978, p. 81) description of the murmur of Mitral Insufficiency (b). Textbook (Moller, 1978, p. 91) description of the murmur of Coarctation of the Aorta (c).

MI and that this may well explain the performance of less experienced subjects. While the experts' interpretations of this murmur surely contributed to their erroneous diagnoses for the case, it is not clear why they interpreted the murmur as they did.

The handling of the Mitral Insufficiency murmur does not explain all of the final diagnoses for Case 5. If one ignores the BAV (Aortic Stenosis) component of the final diagnoses in Table 12, these diagnoses can be partitioned into three basic groups. The first group involves Coarctation only and includes all of the experts and no other subjects. The second group of diagnoses involves Coarctation plus MI and includes subjects S1, S3, S4, T1, and T4. Two subjects (S2, T2) had a third class of diagnosis, involving Coarctation plus an intracardiac shunt (Ventricular Septal Defect). The remaining subject (T3) included both an intracardiac shunt (ASD) and MI in his final diagnosis. The conjoining of shunts with Coarctation in some diagnoses can be used to gain some interesting insights into the dynamics of the diagnosis of this case.

In addition to the murmur already discussed, there are three patient findings presented in the case which might lead a subject to add components, including a shunt, to his main diagnosis of Coarctation. These are the

findings of increased pulmonary vascularity (x-ray), left atrial enlargement (x-ray and EKG) and right ventricular hypertrophy (EKG). These are findings that, from our a priori interpretation of the case, argued strongly for the presence of Mitral Insufficiency, independently of a subject's interpretation of the murmur.

The set of subjects, taken as a whole, raised three basic arguments or interpretations in relation to these three findings. The rationale behind each of these arguments will be outlined below, as a prelude to a final set of results from the case:

1) The pure "obstruction to emptying" argument: a schematic for this argument is given as Figure 12(a) and reference to that figure may help the reader understand the verbal discussion. The pure obstruction argument involves only Coarctation and secondary effects of this condition. Under this argument, the Coarctation makes it difficult for the left ventricle to eject all of its blood, and residual blood "pools" in the left ventricle. Because of pooled blood in the left ventricle, the left atrium has difficulty emptying all of its blood, blood pools in the left atrium and this chamber enlarges. For analogous reasons, blood pools in the pulmonary veins causing pulmonary venous

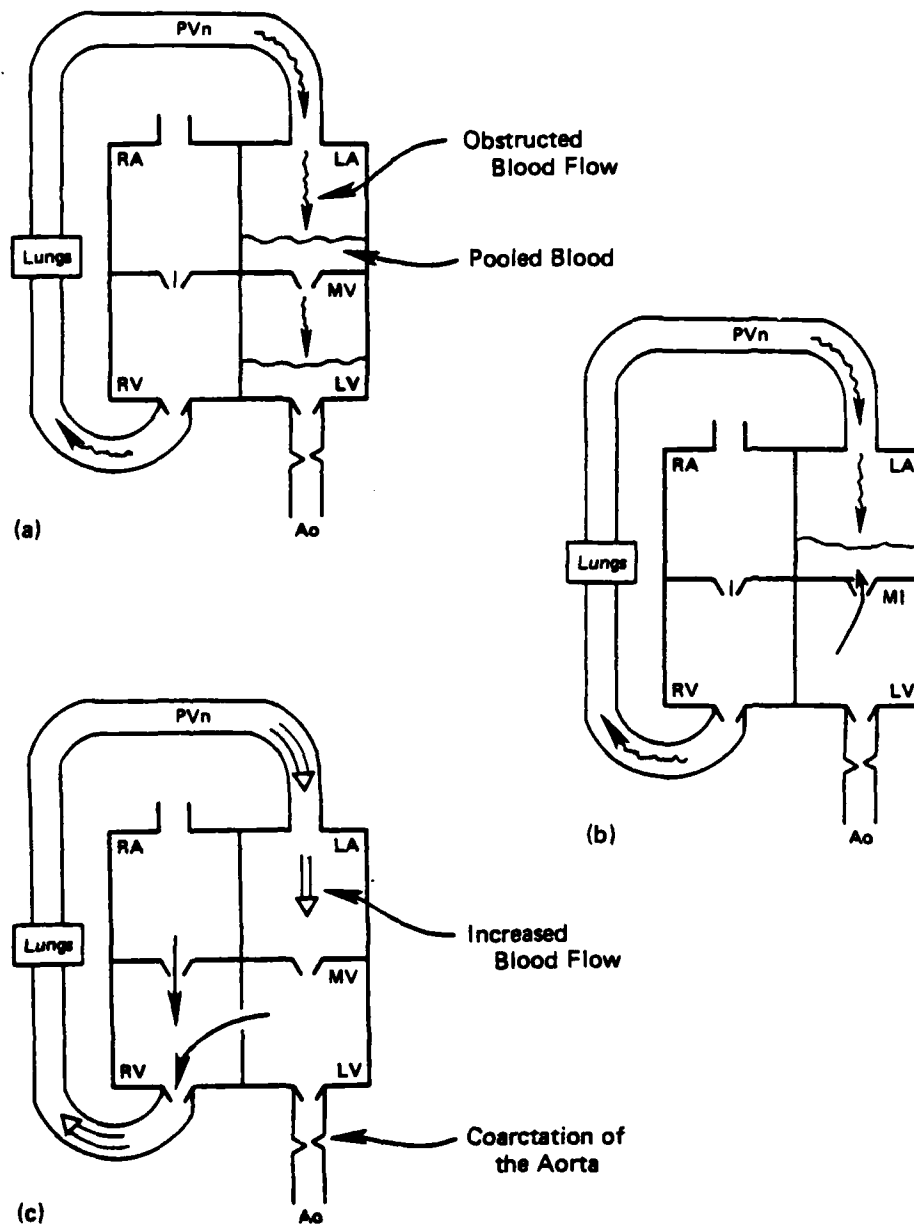


Figure 12. Schematic for the "pure obstruction" argument—Case 5(a). Schematic for the "augmented obstruction" argument (b). Schematic for the "shunt" argument (c).

congestion. The right ventricle, working hard to push blood through this stagnated system, thickens its musculature, that is, hypertrophies.

An example protocol showing this argument applied to left atrial enlargement and prominent pulmonary vasculature (the x-ray, item 22), is given in Figure 13(a). Experts must have relied on something akin to the "pure obstruction to emptying" argument to justify their basic interpretations of the case. While this interpretation is generally reasonable, it appears to involve a commitment to at least a partially "failing" left ventricle (congestive heart failure), that is, to a left ventricle that has dilated and lost the capacity to eject its normal component of blood (Hartmann, Goldring, Strauss, Hernandez, McKnight, and Weldon, 1977, p. 201-204; Moller, Amplatz and Edwards, 1971, p. 49). Substantial signs of congestive heart failure are not present among the patient findings of Case 5.

2) The "augmented obstruction to emptying" argument. A schematic for this argument is given as Figure 12(b). This is the interpretation that includes Mitral Insufficiency with a Coarctation and would lead to the correct diagnosis of the case. The augmented argument

(22) The x-ray shows moderately enlarged heart and left atrial enlargement. The pulmonary vasculature is prominent and perhaps slightly increased.

(a) The pure "*Obstruction to Emptying*" argument

E3: Dilatation of the left ventricle is presumably present. Dilatation of the left atrium is, uh, secondary to elevated end-diastolic pressure in the left ventricle. The prominent pulmonary vasculature, uh, is, uh, most likely related to increased, uh, pulmonary venous markings, uh. These features are consistent with significant left-sided obstructive lesion and would be consistent with a severe coarctation of the aorta.

(b) The "*Augmented Obstruction*" argument

S3: Ok, the moderately enlarged heart and left atrial enlargement ah, that fits with the regurgitation across the A-V (Mitral) valve and the enlarged heart probably could fit with the ah, there probably is hypertrophy. We'll see that on the EKG most likely. Pulmonary vasculature is prominent. Now let's see that ah, I guess that could be due to ah, back flow. Back pressure would be on the venous side so you would look for exactly the characterization of the vasculature.

(c) The "*Shunt*" argument

S2: Ok, the ah, x-ray showing a moderately enlarged heart could be ah, evidence of a ah, coarc, ah, with the increased pressure transmitted to the left ventricle. But, ah, with the left atrial enlargement it would also point to me that there is probably a volume overload. And the fact that the pulmonary vasculature is prominent is ah, very ah, significant, I think, in that you wouldn't expect to find that in a pure coarc. And ah, one would have to be, I think, a little more, bit more suspicious that maybe it may be a combination of lesions with a coarc being in evidence but also the a ah, VSD (Ventricular Septal Defect) which would account for both the murmur that is heard and also the increased flow to the right side and having, and thus giving you the left atrial enlargement and also the increased pulmonary vasculature.

Figure 13. Protocol examples of three basic arguments used to explain Left Atrial Enlargement and other key findings of Case 5.

follows the same fundamental logic as the pure obstruction argument in that its basic ingredient is one of obstruction of inflow to the left side of the heart. Under the augmented interpretation, blood back-flows or regurgitates from the left ventricle through the insufficient mitral valve to the left atrium. The resulting increased blood volume in the left atrium causes the chamber to enlarge and also poses obstruction to emptying of the pulmonary veins. The explanations for pulmonary venous congestion and right ventricular hypertrophy are the same as under the pure obstruction argument.

This argument has all the advantages of the pure obstruction interpretation without requiring an element of congestive heart failure. An example of this argument applied to the x-ray is given in Figure 13(b).

3) The "shunt" argument. A schematic for the shunt interpretation is given in Figure 12(c). This interpretation is fundamentally different from the others. Under this argument the left-sided heart effects under discussion are the result, not of obstruction to inflow of blood to the left side, but rather, result from an abnormally large volume of blood flowing into the left-sided chambers, from the right side of the heart.

According to the shunt argument, some kind of abnormal communication (e.g., a hole in the ventricular septum, VSD, or an open ductus arteriosus, PDA) exists between the left and right-sided heart systems. This communication allows blood to shunt or cross over from the left to the right side during certain parts of the cardiac cycle. During other parts of the cycle, this augmented right-sided blood is pushed through the lungs by the right ventricle into the left-sided chambers of the heart. Right ventricular hypertrophy, increased pulmonary vascularity, and left atrial enlargement are related to this increased volume of blood flow.

A protocol example of this argument applied to the x-ray is given in Figure 13(c). While the shunt explanation is a rather poor explanation for Case 5 taken as a whole (among other things, the murmurs are wrong), a shunt is a reasonable "local" explanation for the target findings under discussion. Subjects who added a shunt to their final Coarctation diagnosis did so under the argument just outlined.

The three arguments just discussed suggested a kind of unanticipated Logical Competitor Set for Case 5, a set of plausible "local" alternatives for a small group of findings in the case. This LCS is reminiscent of the alternative explanations for the finding of

"systolic murmur" in Case 2 reported in Table 6. In the spirit of the investigations of the use of Logical Competitor Sets carried out on other cases of the study, it was decided to investigate application of alternative causal arguments to the three target patient findings discussed above for Case 5 (left atrial enlargement, right ventricular hypertrophy, increased pulmonary vascularity).

Because the "pure" obstruction argument and the "augmented" obstruction argument greatly overlap in their causal dynamics, it proved impossible, in general, to discriminate these arguments in protocols given by subjects (the relatively clear examples of Figure 13 notwithstanding). However, discrimination between the two obstruction arguments and the shunt argument was, in general, straightforward. Hence, it was decided to analyze the application of a general "obstruction" argument (pure or augmented) versus a shunt argument to the target items of Case 5.

Table 14 shows the results of this analysis. An "x" under one of the findings and arguments in this table indicates that a subject applied the argument directly to the interpretation of the finding, regardless of the evaluative "tone" of this interpretation, that is, whether or not the subject judged the argument to

Table 14
Case 5: Two Main Causal Arguments Applied to Target Findings

Argument	Subjects	Finding					
		Increased Vascularity		Left Atrial Enlargement		Right Ventricular Hypertrophy	
		Obstruction	Shunt	Obstruction	Shunt	Obstruction	Shunt
Students	S1		X	X			
	S2	X	X		X		X
	S3	X		X		X	
	S4	X		X		X	
Trainees	T1	X	X	X		X	
	T2	X		X	X		X
	T3	X	X	X	X		X
	T4	X	X	X	X		X
Experts	E1	X	X	X	X	X	
	E2	X	X	X	X		
	E3*	X	X	X		X	
	E4*	X	X	X	X	X	

Note. (X) simply indicates that a subject raised the designated argument in protocol in response to the designated patient finding; it indicates nothing concerning whether the subject judged the argument to be a good or bad one as applied to the finding.
* The two experts with more than twenty years of experience.

yield a good or poor explanation for the finding.

The intent of Table 14 is to determine which subjects raised or gave a "hearing" to the two main alternative explanations for each finding. One way to address this question is to ask which subjects raised both arguments in relation to any one target data item. Only one student (S2) applied both arguments to a common target data item while all trainees and experts meet this minimal criterion for comparative consideration. Once again we see evidence that a consequence of increasing experience is the grouping together of reasonable alternative explanations. Students, in general, appear not to have as many interpretive options available to them. It is as though, for inexperienced subjects, each finding has an idiosyncratically salient and lone causal interpretation. For example, even though subject S2 raised two arguments in response to "increased vascularity", his protocol in response to this finding (Figure 13c) suggests he did not recognize the general merit of the obstruction argument in the interpretation of this finding.

The results of Case 5 have raised several general issues to be discussed in conclusion to this case. The case was originally intended to be a test of whether subjects would add multiple components to a

diagnosis in a case where, from a prior analysis, such addition seemed necessary. In general, less experienced subjects added these components while the most experienced subjects did not.

Another case in the study, Case 2 - TAPVC, showed students who created piecemeal, unintegrated combinations of abnormalities as diagnoses even in a case where these fragments could easily be integrated within a single, more global interpretation. We have also seen, among less experienced subjects, a kind of "local reactivity" to particular data items of a case. Recall the PAPVC diagnoses in Case 2 that seemed to be based largely on a single x-ray finding, the student difficulties with Pulmonary Stenosis on that same case, and the Truncus diagnoses in reaction to the heart sound finding in Case 4. Other cases have shown novice dependence on "textbook" or classic descriptions of patient findings. Since the correct diagnosis for Case 5 is, in a sense, fragmented, and because the components seem to have had classic cues in patient data (e.g., see Figure 11), it would appear that this case "played into the hands" of the novices.

The best explanation for what happened to experts in this case would seem to involve the naturally

occurring rarity of Mitral Insufficiency in combination with the disease Coarctation of the Aorta. In post-experimental discussion, expert E3 claimed never to have seen any patient with this combination, while expert E4 thought he might have seen "less than half a dozen" in over 500 clinical cases of Coarctation. A report of the clinical incidence of various abnormalities in conjunction with Coarctation (Hartmann et. al., 1977, p. 200) does not even list Mitral Insufficiency among the possibilities and the author has yet to find MI in such a listing anywhere. In contrast, BAV and various shunts are fairly common. Given that all the experts were adamant in claiming that they could account for all data of the case under a "pure obstruction" argument and given the rarity of MI in combination with Coarctation, the experts stayed with Coarctation alone. "Never two if one will do" as one expert noted, although the present case shows how this expert tendency toward integration can sometimes lead to trouble.

The expert consultant to the project claimed to have seen a substantial number of patients with a Coarctation-MI combination and this probably accounts for

his consistent correct interpretation of this case. Interestingly, students and trainees had no trouble forming the Coarctation-MI combination based on "text-book" cues for individual components, although it is safe to assume they had no clinical experience with it. This would seem to indicate that with increasing experience, the foundation of a subject's disease knowledge changes from one based primarily on textual materials (e.g., reference manuals or instructional materials) to one weighted more heavily on clinical experience.

Summary

The results from five diagnostic cases were presented in this chapter. Cases were designed to test different aspects of disease knowledge in subjects. Two kinds of results were presented for each case. The first of these involved the use by subjects of a designated set of diseases (in the first four cases, a Logical Competitor Set) as diagnostic hypotheses during the course of the case. The other kind of result was an analysis of key diagnostic errors committed by subjects in their handling of the case.

Case 1. The first case, Subvalvular Aortic Stenosis (SubAS), was designed to test subjects' working knowledge of variants of a disease in addition to a "classic" version, and the grouping of these variants in memory.

All subjects used classic Valvular Aortic Stenosis (ValvAS) as a hypothesis, supporting our pre-supposition that this disease variant would be highly salient in the knowledge base of subjects. Only one student ever utilized both of the other designated disease variants, Subvalvular and Supravalvular Aortic Stenosis (SupAS), and he made only passing reference to

these in his diagnosis. In general, students focused almost exclusively on the classic variant during the course of the case. Half of the trainees used only Valvular Aortic Stenosis in the case, while the other two trainees utilized the full set of variants. No expert was limited to only the classic variant, and three experts (including the two most experienced) used the full set of variants. All subjects who used the full set of variants did so in one of two patterns; they either used all variants together at a data point which represented a strong cue for Valvular Aortic Stenosis or they used all variants together at a data point immediately following strong evidence against ValvAS. These patterns were taken as evidence for memory grouping among the disease variants.

No students, two trainees, and the two most experienced experts diagnosed the case correctly as Subvalvular Aortic Stenosis. All other subjects chose the classic version, Valvular Aortic Stenosis. Six of these subjects (three students, two trainees, and one expert) never generated SubAS as a hypothesis, even though they had generated ValvAS. Among other possible explanations, sparseness (lack of connectedness) in disease knowledge was proposed as contributing to this lack of activation:

ValvAS and SubAS were not strongly enough associated in memory to enable the activation of ValvAS to spread to the other variant.

Two students who diagnosed the case incorrectly were the only subjects who misevaluated a piece of strong evidence against ValvAS, "no systolic ejection click", as confirmatory evidence for this variant. This was interpreted as reflecting imprecision in the students' disease models for the variant. Another student, as well as some of the more experienced subjects who missed the case (including the two less experienced experts) exhibited faulty knowledge of the physiologic cause of systolic ejection clicks. This error in causal knowledge led the subjects to minimize the import of the critical evidence against Valvular Aortic Stenosis.

Case 2. The second case, Total Anomalous Pulmonary Venous Connection (TAPVC), was designed to test subjects' memory grouping of a Logical Competitor Set (LCS) of diseases, a set of diseases with "increased right-sided blood volume." Other diseases in this set include Atrial Septal Defect (ASD), Endocardial Cushion Defects (ECD), and Partial Anomalous Pulmonary Venous Connection (PAPVC). Of these ASD was assumed to be classically associated with the strong hypothesis

"triggering" cues of the case.

All subjects generated Atrial Septal Defect as a hypothesis, supporting our presupposition that this disease is strongly associated with a set of cues in the case. No student utilized the full set of competitor diseases as hypotheses in addition to ASD. One trainee utilized the full LCS as did the two most highly experienced experts, E3 and E4. The two most experienced experts used the full LCS according to the same patterns in which they had used the full set of competitors in Case 1. Expert E3 used most of the competitor set diseases together at the data point of first strong evidence for ASD. Expert E4 used only ASD at this point, but expanded to the full LCS at a later point of strong disconfirmatory evidence for ASD. The trainee who used the full LCS did so in a pattern similar to expert E3. As in Case 1, the use of the full LCS in these patterns was taken as evidence for memory grouping or categorization of the competitor set diseases.

Only four subjects, two trainees and the two high-level experts, diagnosed the case correctly as TAPVC. Again, a number of subjects (three students, one trainee, one expert) never generated the correct hypothesis, partly accounting for their erroneous

diagnoses of the case.

Among errors of commission, two students had final diagnoses outside of the Logical Competitor Set of diseases. These diagnoses involved combinations of multiple diseases, and the primary locus of error for these students was traced to limitation in knowledge of alternative causes for a particular systolic murmur; the students, unlike more experienced subjects, considered only one of two feasible causes for this finding - the wrong one for the context of the case - and this interpretation encouraged multiple diseases in a diagnosis. Another student diagnosed the case as Endocardial Cushion Defect, and seriously misinterpreted the EKG axis in relation to this disease. This was taken as reflecting error or imprecision in this subject's disease model for ECD. One of the trainees demonstrated overrestrictiveness (imprecision) in her allowable age range for TAPVC, causing her inappropriately to reject this disease in favor of ASD. Finally, four subjects, spanning the dimension of experience, diagnosed the case as PAPVC. These subjects were strongly influenced by a classic cue for PAPVC on x-ray. When this cue elicited PAPVC, it did not, in turn, elicit TAPVC or other LCS members, again suggesting sparseness (lack of

connectedness) in disease knowledge.

Case 3. The third case of the study, Patent Ductus Arteriosus (PDA), was designed to test subjects' grouping of a Logical Competitor Set of diseases in a case where there was an early, classic cue for the correct disease and no subsequent data were discrepant with this interpretation. The case was different from the first two cases in that in those cases, the first strongly induced disease interpretation was wrong. Because determining the correct diagnosis was so straightforward, Case 3 was seen as a severe test for hypothesis grouping. PDA and the other four diseases in the LCS for Case 3 constituted a physiologic category of "extra-cardiac shunts" and related conditions.

All subjects generated PDA in the case as expected, and all subjects concluded this disease as a final diagnosis. One student, three trainees, and high-level expert E4, generated only PDA among LCS members during the course of the case. All other subjects generated at least one other LCS member in addition to PDA. Only one subject, high-level expert E3, used all five members of the LCS during the case, and he did so in a manner consistent with his complete use of the competitor sets in other cases. That is, he generated

the entire set as competing hypotheses in close proximity to the classic cue for Patent Ductus. Another expert, E2, considered most (3 of 5) of the LCS, and he did so in a pattern similar to E3; no non-expert subject utilized as much of the LCS. The performance of high-level expert E4 was seen as consistent with his consideration of the full LCS in other cases, in that in the first two cases he had expanded to the full LCS only after receiving disconfirmatory evidence for a preferred member; in Case 3 there was no such evidence.

Case 4. The fourth case, Pulmonary Atresia (PAT), was, again, designed as a test of clustering in memory, by subjects, of a physiologic category of diseases - "cyanotic diseases with decreased pulmonary blood flow." This Logical Competitor Set for Case 4 included Ebstein's Malformation of the Tricuspid Valve (Ebstein's) and Tricuspid Atresia (TAT) in addition to the correct disease, Pulmonary Atresia. No member of the LCS was seen as generally more prototypic for the category than the others.

No student utilized the entire Logical Competitor Set for the case, and two students generated only one member. No other subjects generated this few. In addition, two students were the only subjects who never

created the correct disease, Pulmonary Atresia, as a hypothesis. In contrast to the students, all four trainees and three experts (including high-level expert E3) used the full competitor set as hypotheses. However, the experts and trainees utilized the full competitor set according to different patterns. The experts, unlike the trainees, used all LCS members together at the point of the case where the category "cyanotic disease with decreased pulmonary blood flow" could first be established. No trainee used the full set at this point, suggesting that the LCS diseases did not constitute a functional diagnostic category for these subjects - at least not to the same extent as for the experts.

Highly experienced expert E3 utilized the full LCS in Case 4 in the same way he had done in other cases, that is, he utilized all member diseases at the first clear opportunity to do so. While expert E4 named only the correct disease, Pulmonary Atresia, at this point, he also articulated the category name for the LCS diseases at this point and his performance was interpreted as being consistent with his handling of LCS diseases in other cases - since no subsequent data were incompatible with the particular LCS member he chose for focus.

Three experts diagnosed the case correctly as Pulmonary Atresia while this final diagnosis was offered by only one student and one trainee. Two general types of error were demonstrated by subjects who missed the case. Two students diagnosed the case as Truncus Arteriosus, and a particular data item, "single second heart sound", was proposed as the most important source of this misdiagnosis. This finding is consistent with all LCS members but is perhaps most clearly representative of Truncus. Students' handling of this data item was interpreted as reflecting a kind of classical dependency in the knowledge base. Most trainees diagnosed the case as Tricuspid Atresia and the main source for this misdiagnosis was traced to an overly liberal range of expected values for the EKG axis under this disease - imprecision in subjects' disease models for Tricuspid Atresia.

Case 5. The final case of the study was characteristically different from the others in that there was no issue of Logical Competitor Sets. The case contained a pathognomonic cue for the main disease in the diagnosis, Coarctation of the Aorta (Coarc). The case was designed to test whether subjects would detect the need for adding two additional diseases, Bicuspid Aortic

Valve (BAV) and Mitral Insufficiency (MI), to a correct diagnosis.

In general, the results of the case were opposite to what was expected. Most significantly, no expert added Mitral Insufficiency to his Coarctation diagnosis while many less experienced subjects did. The success of the less experienced subjects in adding this component was attributed to a rather "textbook" description of the murmur of MI which was presented as a data item in the case. The failure of the experts was attributed to the clinical rarity of MI in combination with Coarc and to a general tendency among these subjects to avoid multiple-disease diagnoses if at all possible. The argument by which experts judged Mitral Insufficiency to be unnecessary was sketched. Case 5 is an example of a situation where expert tendency toward integration may lead to difficulty.

An unanticipated analysis for Case 5 involved subjects' use of two main physiologic arguments, "obstruction" and "shunt", for a set of important patient data items in the case. Students generally seemed limited to one or the other of these interpretations while more experienced subjects generally applied both. This restrictiveness in options for interpretation was

consistent with novice performance in similar situations in other cases.

Based on the performance of the two highly experienced experts (E3, E4) on the cases of this study, an expert form and an expert substance for diagnosis can be proposed. The expert form involves the full, active use of a set of physiologically similar diseases (the Logical Competitor Set) in a case, diseases that have similar clinical presentations. Expert E3 used the full LCS in all four cases where this was tested - more than any other subject. Expert E4 explicitly used the full LCS in two cases and utilized a category hypothesis for the LCS diseases in another. The use of the Logical Competitor Set by experts, generally in close proximity to the most important patient findings associated with members of the set, is interpreted here as evidence that these diseases constitute a unit or category in memory.¹³ Since diseases in the Logical Competitor Set are likely to be confused with each other, in the "long-run" it would seem highly adaptive

¹³

There are more direct ways of assessing this claim, such as experimental tasks that require subjects to draw on their memories in ways that allow measurement of inter-response times (e.g., Reitman, 1976). Studies of this sort are currently under planning.

for a diagnostician to be readily able to engage other members of the set in contexts where there is strong reason to believe one of them is a good candidate for a case. Memory "unitization" (c. f. Anderson, 1980) of LCS diseases in some sort of category-like structure would provide this capability to the expert.

Expert substance in diagnosis refers to correct data interpretations and data evaluations, in relation to the Logical Competitor Set of diseases, necessary to isolate the correct member. Expert E4 diagnosed four of the five cases correctly, more than any other subject. He missed only the controversial Case 5. Expert E3 diagnosed three of five cases correctly, as many as any other subject. Isolation of the correct LCS member is supported by precision in the clinical data expectations contained within individual disease models.

For the two high-level experts in the study, two distinct methods of utilizing the LCS were also identified: (1) Precaution. This involves the generation and use together as hypotheses of the full set of logical competitors, enabling them to be weighed against each other and data. (2) Extraction. This method involves more aggressive focus on a member of the set, with full expansion to the remainder of the set should

disconfirmatory evidence for the target member be found.

The two less experienced "experts", E1 and E2, performed much like trainees on the cases of this study. In particular, each diagnosed only two cases correctly, a figure eclipsed by three trainees. E2 used the full LCS in only one case, while E1 did so in two cases. In addition, each showed interpretive errors similar to those made by less experienced subjects, in particular the handling of the "click" in Case 1, and the susceptibility to classic cues in Case 2. These subjects did enable us to see that the development of high-level skill in pediatric cardiology takes a long time.

Medical students after six weeks of training and clinical practice in the field represented by the cases, generally showed neither expert form nor substance. Students almost never considered the full LCS (one student on one case) and focused on the "classic" members in cases which encouraged this. This suggests that LCS members, when they exist in memory at all, are represented in a more isolated form than for experts. Two students diagnosed no case other than Patent Ductus correctly, while the other two diagnosed one disease in addition to PDA. Errors of interpretation and evaluation (shared with intermediate level subjects)

included several basic types: (1) Mundane factual errors. These are just factual errors about which findings "go with" which diseases. (2) Causal errors. These are errors concerning how observable data are related to underlying physiology. (3) Imprecise tests. These are either overly general or overly restrictive tolerances on the range of variability allowed in an expected clinical finding for a disease. (4) Interpretive restrictiveness. This refers to restriction in the number of interpretations that are applied to a finding. These errors are interpreted as reflecting imprecision in subjects' models for diseases.

The "trainees" in the study (residents in pediatric cardiology at the end of their training and first-year fellows in pediatric cardiology) at times showed performances that looked very much expert-like and at times could not be distinguished from the students. The number of trainees in each case who used the full LCS fell between the students and experts on two cases (Case 1 and Case 2), was the same as the students on Case 3, and actually exceeded the experts on Case 4. Three trainees diagnosed three cases correctly - on a par in this regard with high-level expert E3. Depending on subject and case, trainees at times exhibited

the types of errors discussed above for the students. The ultimate diagnoses of the trainees, unlike those of the students, were always at least within the LCS, if not correct. This suggests that for trainees who missed cases, the main problems involved lack of connectedness in memory among LCS members, or imprecision in knowledge necessary for discriminating LCS members correctly.

4. DISCUSSION

The study demonstrates that diagnosticians' "disease knowledge," a memory store of disease models and the memory organization among them, is important to successful diagnosis and does explain differences between expert and less expert performance.

The general "process" of diagnosis, although this has not yet been directly addressed from the study, seemed similar for all subjects; that is, all subjects generated diagnostic hypotheses at various times in cases and proceeded to test these against subsequent data. The major differences among subjects which have emerged so far concern their handling of a set of "good moves," that is, the Logical Competitor Sets. More experienced subjects tend to consider them in groups, and evaluate them correctly.

The study did not set out to show that highly experienced practitioners diagnose differently from novices; this should go without saying. The intent was to learn something about the medical knowledge that diagnosticians use, the way this influences performance, and the ways this knowledge changes as people acquire experience in a field. Medical students, after only six weeks of training in the field of interest, were included as subjects because these individuals represent

the "starting point" in a long learning process.

The Nature of Knowledge Change

What has been learned about the nature of knowledge change? It seems clear that the whole learning process starts with a small set of "classic" training concepts where these include particular diseases, descriptions of expected patient findings under these diseases, and rules for disambiguating diseases in this starting set. The learning of these training concepts is encouraged by the selection of content for inclusion in introductory training materials, that is, introductory textbooks and classroom instruction. The diseases are the common ones or the clearest instances of a physiologic type, the patient data descriptions are "prototypic" or average, and the rules of evaluation are overly simplified. We have seen several instances where the locus of novice errors could be traced fairly directly to such statements in the introductory textbooks to which the subjects had been exposed.

One might wonder why the introductory materials are as they are. There are probably several good reasons. First, it is not clear that the "ultimate textbook", one that documents all of the context-sensitive nuances of patient presentation and all of

the "exceptions" to the rules of evaluation could even be written. Medical research and the clinical case documentation literature periodically change what such a book would contain. Even if the ultimate "book" could be written, it would probably be a poor introductory textbook. It would be such a morass of detail that students might learn nothing at all. Secondly, there are sound cost-benefit reasons for emphasizing the common things. If a new practitioner cannot be taught everything, it is obviously right to give him the tools to work with the problems he is most likely to encounter. Most students who take the introductory course in pediatric cardiology do not even become pediatric cardiologists. The training objective for these individuals is probably only that they be able to recognize problems for referral. Finally, for those who specialize, a long apprenticeship is assumed. A minimum of six-seven years of clinical and other training would be required before the students in this study could be certified as pediatric cardiologists. What is important is not that they be taught everything but rather that they be given the cognitive "anchorage points" to enable them to benefit from the experience to follow.

With experience, the practitioner is exposed to and adds to memory additional disease models beyond the introductory set. The expert's "large vocabulary" of discriminable instances is now well documented (e.g., Chase and Simon, 1973a). Concurrently with the simple addition of disease models to memory, there is an embellishment of the compositional features of a disease that are encoded within each disease model. These are features representing the diseases' internal physiology and clinical presentation. The expert simply knows more defining attributes of a disease. Rosch (Rosch and Mervis, 1975) reports an intriguing anecdote that bears on this issue. One of her standard paradigms asks subjects to list as many features as they can of various common objects, e.g., an airplane. Among ordinary college subjects the number of features listed for any particular object is fairly stable. However, one subject was found to list many times more features for an airplane than average; on follow-up he was discovered to be an airplane mechanic. In some of our own work, we have found that expert physics problem solvers actively use "transformed" or "abstracted" features of a physics problem statement that novices do not even seem to recognize (Chi, Feltovich, and Glaser, 1980). In Case 1 of the present study, there were some inexperienced subjects

who did not generate any Aortic Stenosis hypothesis until after the presentation of the critical finding of "no click". The fact that they did not return to this finding after the Valvular Aortic Stenosis model was engaged, suggests that this feature was not represented in their model of the disease; they may have had no expectation at all regarding a click. Recall that in Case 2 of the present study some inexperienced subjects seemed to view the Pulmonary Stenosis issue (Table 6) as involving only one dimension, that is, "orifice size", when in fact the problem involves the two interacting feature dimensions of size and flow. This is highly reminiscent of the "dimensional restrictiveness" or paucity of encoded problem features reported by Siegler (Siegler, 1976; Siegler, 1978) for inexperienced problem solvers.

As an individual encodes more features of a disease, this provides opportunity for differentiating the disease into subtypes (c.f., Anderson et.al., 1979). As an illustration, if a person encodes only the features of "height" and "weight" for people, he is quite limited in the discriminations he can make among people. It is clear that the disease knowledge of a highly experienced diagnostician is highly differentiated within a disease type. In the present study the case explicitly

designed to assess this was Case 1, where the increasing differentiation was demonstrated. It can be noted that for Case 2, TAPVC, high-level expert E3 raised and considered no fewer than nine different varieties of TAPVC where each of these was distinguished by slight anatomical difference.

The differentiation of disease knowledge aids the development of precision in the clinical expectations associated with any particular disease model. If possible distinctions among versions of a disease are not made; that is, if they are in a sense all seen as the same thing, then the associated variability in clinical manifestations among patients will be great. However, when an expert represents in memory, say, nine different anatomic versions of TAPVC with each of these perhaps differentiated into more specific versions by severity and age of presentation in a child, then the clinical expectations associated with each of these "micro models" can be highly specific.

Precise clinical expectations, in turn, contribute to precise rules of evaluation for patient data. This is the difference between the "left axis deviation" rule used by less experienced subjects in Case 4 and the experts' "down around zero or slightly negative" rule used in evaluating the EKG axis in that case with

respect to Tricuspid Atresia (See Table 10 and the discussion around it). Again, in Case 1, one can see a nice example of how differentiation of a disease contributes to correct evaluation. In his protocol given in Figure 4, expert E3 raises the one micro version of Valvular Aortic Stenosis in which a "click" is not expected. This is the version with a pressure gradient between the left ventricle and aorta (over the valve) of greater than 100 mm, that is, "Aortic Stenosis of a very severe degree." Under this version, other data of the case would have been different from those presented. The expert was able to bring the appropriate (i.e., moderate severity) version of Valvular Aortic Stenosis to bear on the evaluation, and to reject this version.

The embellishment of the feature set within disease models aids generalization as well as discrimination. Every additional feature represented for a disease is a potential feature of similarity with another disease; hence, the potential of generalization to a category of "diseases that share feature x" exists (c.f. Anderson et.al., 1979). The LCS analyses throughout this paper are taken as evidence that such groupings are pervasive in the more experienced knowledge base. Tight memory organization among competitor

diseases, in a category (Cohen, 1966) or similar type of memory unit (Anderson, 1980), supports diagnosis by providing interdisease activation; when one member is activated, other plausible candidates are likely to be considered.

Students learn some disease groupings directly (e.g., Moller, 1978, p. 46). These, analogously to other teaching concepts, might be thought of as a set of "starting-point" disease categories. With experience and embellishment of feature sets, a diagnostician augments this initial category set, often creating useful categories that "cross over" the original "classic" set. Case 2 from the present study is a good example. One might wonder how it is that a number of subjects on this case could generate and extensively consider the hypothesis "Partial Anomalous Pulmonary Venous Connection" and never once even think of the correct disease, "Total Anomalous Pulmonary Venous Connection", a disease that even in its name is so similar. In the classic categorization of diseases, PAPVC, ASD, and ECD, three members of the Logical Competitor Set for this case, all go together in a category of "acyanotic heart diseases" (See Figure 14), while the final LCS member, TAPVC, is in a different category, "cyanotic heart

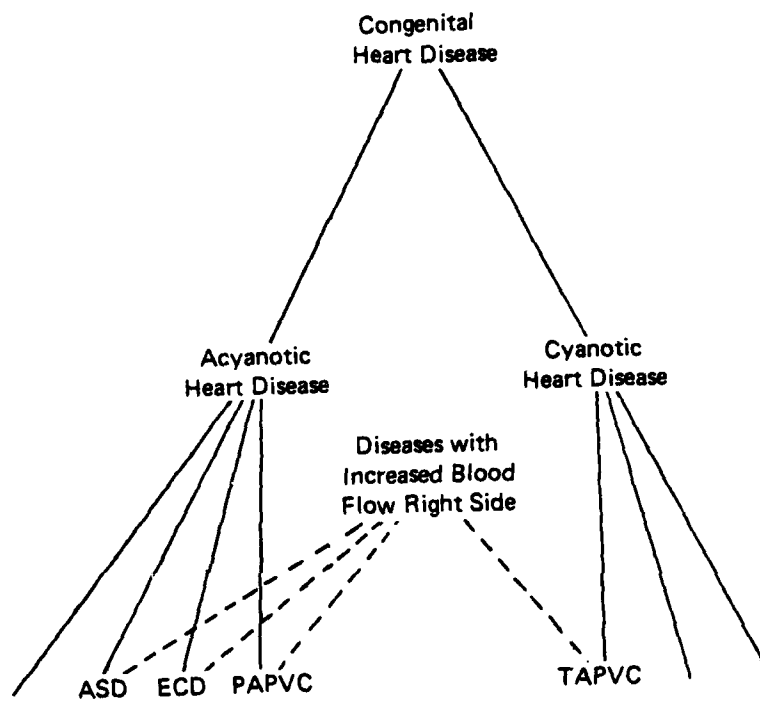


Figure 14. The classic categorization (solid lines) of the members of the Logical Competitor Set, Case 2, and expert regrouping (dashed lines) of these diseases.

disease." One explanation for these subjects is that they became "stuck in a chunk"; that is, they were in the wrong branch of their classic hierarchy and were not able to benefit from associative hypothesis triggering or "spreading" activation (e.g., Anderson, 1976). The two high-level experts, on the other hand, had created a category for the LCS members that crosses the classic categorization scheme (See Figure 14). Creation of this category required them to represent a new disease feature, the feature of "increased blood flow on the right side."

Our speculation is that many kinds of logical and practical groupings exist for the expert, tailored to different problem contexts and even different phases of patient data collection, e.g., "the not too sick two-day-old child" in the very early phases of diagnosis. The totality of these groupings for the expert need not be strictly hierarchical; that is, the groupings may "cross over" each other in many different ways forming more a "lattice" structure than a formal hierarchy (Pople, 1977).

The pervasiveness of memory groupings in the expert is a logical extension of the general "perceptual chunking hypothesis" of Simon and Chase (Simon and

Chase, 1973) and all of its ramifications (e.g., Chase and Chi, 1980). The cognitive "chunks" for an environment that people create with experience are those that serve their goals for functioning in that environment (e.g., see Egan and Schwartz, 1979, for "electronics trouble shooters") and, in addition to activational functions cited above, probably serve the problem solver by focusing problem solving activity as we will now discuss.

The Relationship Between Knowledge and Problem Solving

As problem solving research has moved from semantically "lean" domains, e.g., various toy problems such as the "Tower of Hanoi" and "cryptarithmic" (e.g., Newell and Simon, 1972), to semantically rich domains such as physics or "engineering thermodynamics", the importance to problem solving of domain knowledge has been increasingly recognized (e.g., Bhaskar and Simon, 1977).

One of the issues we set out to address with this study was the relationship between domain knowledge and general "problem solving processes." One way to address this issue is from a framework for problem solving processes set out by Newell (Newell, 1969). Newell proposed a power-generality dimension for problem

solving procedures. General procedures are those that apply widely (the conditions for their applicability are non-stringent) but offer little guarantee of success. Examples are "means-ends" analysis and "hill climbing." Powerful procedures are those that have well specified conditions which must be met for their applicability, hence, are tailored to particular closed environments. An example is the formula for solving quadratic equations. Our work and that of others (e.g., Elstein, et. al., 1978) has shown that the general problem solving procedure for diagnosis is one of hypothetico-deduction and that all subjects, regardless of experience, share this general approach. However, the present study has shown that this alone "will not get one very far". The general process must be backed up by a rich body of accurate, well organized medical knowledge.

With regard to procedures for diagnosis, we speculate that as disease knowledge develops to include numerous pragmatic disease clusterings, as outlined earlier, corresponding sets of relatively powerful procedures, in Newell's sense, are concurrently created. Hence, we would propose that as the diagnostician establishes various partitions of the disease space, for example, the Logical Competitor Sets of various kinds,

he also establishes associated strong "local" procedures for working within those regions of the space. This would mean, for instance, that the experienced diagnostician would have relatively intact or readily assembled "plans" (Sacerdoti, 1977; Vahnlehn and Brown, 1979) or "scripts" (Schank and Abelson, 1977) for discriminating among hypotheses within conceptual disease groupings of various kinds and levels of generality.

The worth of any such plan or script depends on a great deal of specific content of knowledge, including the best alternative disease candidates, precise clinical data expectations for each alternative disease, and differential patient data items that can be used to distinguish among the set of target disease candidates.

Novices in the present study at times showed deficiency both in assembling a good set of alternative disease candidates, and in the specific knowledge necessary to test a disease or disambiguate among alternatives. What is happening when less experienced subjects fail to consider a good set of disease candidates or evaluate data items poorly? One explanation is that necessary knowledge is stored in memory incorrectly or not at all (i.e., knowledge "voids" - the person does not know about Subvalvular Aortic Stenosis). Another explanation concerns problems of access; subjects simply

do not retrieve useful knowledge or retrieve it in some faulty manner.

Post-experimental discussions (conducted after all cases were completed) with the subjects from this study indicated that many subjects, when they failed to generate a good set of candidate hypotheses or interpreted items poorly (e.g., the "click" in Case 1), "knew better" in some sense. Under conditions outside the diagnostic task, they could often discuss Aortic Stenosis variants, e.g., Subvalvular Aortic Stenosis, or the import of the "click" in Valvular Aortic Stenosis, etc. One subject called the experimenter on the day after his session, in which he had erroneously diagnosed Case 2 and never thought of the correct diagnosis, to tell him that the correct diagnosis had "dawned on him in the shower."

Psychology has long known that the ability to access and use knowledge that one "has" is situationally dependent (e.g., Melton, 1963); Tulving and Pearlstone, 1966). For example, knowledge medical subjects might display on a "paper and pencil" test is not necessarily what they could display "on-line" in the diagnostic setting. (It was for this reason that the current study, despite its interest in "knowledge," was conducted in a diagnostic context rather than in some other manner). Yet, it is this "task-accessible" knowledge that is

crucial to successful performance. Knowledge must be utilized appropriately in particular contexts where it is needed. The present work and other recent investigations of novice problem solvers (Polson, Atwood, Jeffries, and Turner, in press), suggest that deficiencies of situational access to extant knowledge rival difficulties associated with knowledge error or absolute lack of knowledge for these individuals.

Implications for Training and Assessment of Competence

Some suggestions for helping a new diagnostician develop an appropriate and, especially, a situationally useable knowledge base can be proposed. First, a disease and other diseases likely to be confused with it in a diagnostic setting should be emphasized together in instruction and, to the extent possible, in the clinical experiences of the diagnostician in training. This encourages the memory unitization of these diseases in categories or other kinds of connected knowledge organizations. Unitization aids access and is a hedge against oversight since information in a unit has two modes of "on-line" activation, associations from external events and activations directed by the unit itself (Anderson, 1980; Cohen, 1966). The competitor set category also provides focus for the creation of a plan

for disambiguating among the member diseases. Because real clinical experiences are constrained by the distribution of patients in the training setting, simulated diagnostic encounters (e.g., McGuire and Soloman, 1971; Swanson, Feltovich, and Johnson, 1977) could provide a vehicle for augmenting natural experiences and for packaging pre-specified sets of experiences, for example, the presentation in close proximity of cases involving competitor set diseases. Secondly, tutorial instruction in the diagnostic process itself must attempt to interact with the on-line thought processes of the learner as he engages in diagnostic-like tasks. This is to help ensure that what is to be taught will be connected both to the situational cues and to the state of active memory likely to exist at some later time when the new material will be needed during a real diagnostic encounter (c.f. Flexser and Tulving, 1978; Tulving, 1976). Expert-based instructional devices (computer assisted instruction or decision support systems) that contain expert knowledge and are capable of performing diagnosis in an expert-like manner, could provide diagnostic practice exercises in which the device diagnoses a case in parallel with a "student," prompting alternative hypotheses when they are overlooked, correcting erroneous interpretations, and

offering instruction when this seems necessary (Brown and Burton, 1975; Clancey, 1979; Clancey, Shortliffe, and Buchanan, 1980; Johnson, Severance, and Feltovich, 1979; Swanson, Feltovich and Johnson, 1977). Finally, it would be advantageous if much of the learning of medical content for those in training could be tied as closely as possible to its conditions of ultimate use. The program in medical education at McMaster University (Barrows and Tamblyn, in press) and the problem-based program at Michigan State (Elstein, et.al., 1978), seem prototypic of such an approach. Under these programs, much of the basic medical subject matter (e.g., physiology) that a student learns is organized within representative professional problems, including diagnoses. The problem directs what is to be learned.

In addition to some suggestions for instruction and training, the current research carries a suggestion for the evaluation of clinical competence. While the recent trend in assessment is toward testing clinical skill using diagnostic and other problem solving tasks in addition to more standard paper and pencil testing, most of these attempts, for example, the "patient management problems" of the National Board of Medical Examiners, concentrate on the process of data collection.

They focus on what the testee chooses to "see", that is, questions asked of the patient and lab tests ordered, etc. While this issue is surely important, the present study has shown that it is equally important to consider the inferences a person makes from the data he does see, e.g., the interpretations of basic findings, the hypotheses considered, and the hypothesis evaluations made. Evaluation of the quality of interpretation need not be limited to final outcomes, that is, whether or not the person gets the problem "right". We have shown that careful task analyses of problems can yield intermediate forms, e.g., the handling of Logical Competitor Sets, to which a person's performance can be compared.

Directions for Future Research

Several directions for future research are suggested by the current work. The first of these is the problem of knowledge access and knowledge use. Not much is currently known about the structure of the knowledge base in memory that facilitates its situational use. Yet, this is clearly a critical issue in problem solving within semantically rich domains. A second important focus is to investigate the "local" plans or "scripts" that competent diagnosticians associate with the various partitions of the "disease space" that they

recognize. These partitions and associated plans might range in generality from particular disease categories to highly general definitions of the problem context - for example, "the healthy-appearing five year old." This kind of investigation appears to be the most promising avenue for studying the "procedures" of diagnosis which have hitherto been studied only at their most general level, that is, at the level of hypothetico-deduction. Work of this sort will require a better mapping of the partitions good diagnosticians use - where the current study is only a start. Finally, the current study can be viewed as one step in a cyclical research paradigm that involves experimentation and more formal cognitive simulation. The current Minnesota diagnostic simulation model (Swanson, 1978; Swanson et. al., 1979) was originally designed as a model of the expert and its initial version was built based on studies similar to the present one. As a result of the present study, adjustments and additions to the initial expert simulation model have been made. In addition, as a result of the study, the framework now exists for the creation of a more novice simulation. This may enable the formal study of learning mechanisms (e.g., Anderson, Kline, and Beasley, 1979)

responsible for the transition from "noviceness" to expertise. The simulations will also direct a new cycle of more focused experimentation.

It is hoped that the present study provides some guidance for the study of problem solving in semantically rich domains. Such work requires both task-environment and knowledge-base analysis and the creation of problem solving environments that make the interaction between the problem's information structure and the solver's knowledge structure comprehensible to the observer.

BIBLIOGRAPHY

- Akin, O. Models of architectural knowledge. London: Pion, 1980.
- Anderson, J. R. Language, memory, and thought. Hillsdale, N. J.: Lawrence Erlbaum Associates, 1976.
- Anderson, J. R. Concepts, propositions, and schemata: What are the cognitive units? Paper presented at the Nebraska Symposium on Motivation, 1980. To appear in the proceedings, 1981.
- Anderson, J. R., Kline, P. J., & Beasley, C. M. A general learning theory and its application to schema abstraction. In G. H. Bower (Ed.), The psychology of learning and motivation. New York: Academic Press, 1979.
- Ausubel, D. P. Some psychological aspects of the structure of knowledge. In S. Elam (Ed.), Education and the structure of knowledge. Chicago: Rand McNally, 1964.
- Barnett, J. A., & Bernstein, M. I. Knowledge-based systems: A tutorial. Technical Report TM-L-5903/000/00 (NTIS: AD/A-044883), Santa Monica, Cal.: System Development Corporation, June, 1977.
- Barrows, H. S. Overview of medical problem solving. Paper presented at the Conference on the Role of Problem Solving in Medicine, Smuggler's Notch, Jeffersonville, Vt., October, 1979.
- Barrows, H. S., Feightner, J. W., Neufeld, V. R., & Norman, G. R. Analysis of the clinical methods of medical students and physicians. Final report, Ontario Department of Health Grants ODH-PR-273 and ODH-DM-226, Hamilton, Ontario, Canada: McMaster University, March, 1978.
- Barrows, H. S., & Tamblyn, R. M. Problem based learning: An approach to medical education. New York: Springer, in press.

- Beach, L. R. Cue probabilism and inference behavior. Psychological Monographs, 1964, 78, 1-20.
- Bhaskar, R., & Simon, H. A. Problem solving in semantically rich domains: An example from engineering thermodynamics. Cognitive Science, 1977, 1, 193-215.
- Bobrow, D. G., & Norman, D. A. Some principles of memory schemata. In D. G. Bobrow & A. Collins (Eds.), Representation and understanding: Studies in cognitive science. New York: Academic Press, 1975.
- Brown, J. S., & Burton, R. R. Multiple representations of knowledge for tutorial reasoning. In D. G. Bobrow & A. Collins (Eds.), Representation and understanding: Studies in cognitive science. New York: Academic Press, 1975.
- Brown, J. S., & Burton, R. R. Diagnostic models for procedural bugs in basic mathematical skills. Cognitive Science, 1978, 2, 155-192.
- Buchanan, B. G., & Feigenbaum, E. A. DENDRAL and Meta-DENDRAL: Their applications dimensions. Artificial Intelligence, 1978, 11.
- Carey, S. Cognitive competence. In D. Connolley & J. Bruner (Eds.), The growth of competence. New York: Academic Press, 1973.
- Charness, N. Components of skill in bridge. Canadian Journal of Psychology, 1979, 33, 1-50.
- Chase, W. G., & Chi, M. T. H. Cognitive skill: Implications for spatial skill in large-scale environments. In J. Harvey (Ed.), Cognition, social behavior, and the environment. Potomac, Md.: Lawrence Erlbaum Associates, 1980.
- Chase, W. G., & Simon, H. A. Perception in chess. Cognitive Psychology, 1973, 4, 55-81. (a)
- Chase, W. G., & Simon, H. A. The mind's eye in chess. In W. G. Chase (Ed.), Visual information processing. New York: Academic Press, 1973. (b)

- Chi, M. T. H. Knowledge structures and memory development. In R. S. Siegler (Ed.), Children's Thinking: What develops? Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978.
- Chi, M. T. H. Exploring a child's knowledge of dinosaurs: A Case study. Paper presented at the Society for Research in Child Development, March 1979.
- Chi, M. T. H., Feltovich, P. J., & Glaser, R. Representation of physics knowledge by experts and novices. Technical Report No. 2, ONR Contract No. N00014-78-C-0375, NR157-421. Pittsburgh, Pa.: Learning Research and Development Center, University of Pittsburgh, 1980.
- Chiesi, H. L., Spilich, G. J., & Voss, J. F. Acquisition of domain-related information in relation to high and low domain knowledge. Journal of Verbal Learning and Verbal Behavior, 1979, 18, 257-274.
- Clancey, W. J. Tutoring rules for guiding a case method dialogue. International Journal of Man-Machine Studies, 1979, 11, 25-49.
- Clancey, W. J., Shortliffe, E. H., & Buchanan, B. G. Intelligent computer-aided instruction for medical diagnosis. Proceedings of 3rd Annual Symposium of Computer Application in Medical Care, 1980, 175-183.
- Cohen, B. H. Some or none characteristics of coding behavior. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 182-187.
- Corsale, K., & Gitomer, D. Developmental and Individual differences in mathematical aptitude. Paper presented at the 20th annual meeting of the Psychonomic Society, Phoenix, Arizona, 1979.
- de Groot, A. D. Thought and choice in chess. New York: Basic Books Inc., 1965.

- Egan, D. E., & Schwartz, B. J. Chunking in recall of symbolic drawings. Memory and Cognition, 1979, 7, 149-158.
- Eisenstadt, M., & Kareev, Y. Aspects of human problem solving: The use of internal representations. In D. A. Norman and D. E. Rumelhart (Eds.), Exploration in Cognition. San Francisco: W. H. Freeman & Co., 1975.
- Elstein, A. S., Loupe, M. J., & Erdmann, J. G. An experimental study of medical diagnostic thinking. Journal of Structural Learning, 1971, 2, 45-53.
- Elstein, A. S., Shulman, L. S., & Sprafka, S. A. Medical problem solving. Cambridge, Mass.: Harvard University Press, 1978.
- Ernst, G. & Newell, A. GPS: A case-study in generality and problem solving. New York: Academic Press, 1969.
- Feigenbaum, E. A. The art of artificial intelligence: Themes and case studies of knowledge engineering. Proceedings IJCAI-5, 1977, 1014-1029.
- Fikes, R. A heuristic program for solving problems stated as non-deterministic procedures. Ph.D. Thesis, Carnegie-Mellon University, 1969.
- Fitts, P. M. & Posner, M. I. Human performance. Belmont, Cal.: Brooks Cole, 1967.
- Flexser, A. J., & Tulving, E. Retrieval independence in recognition and recall. Psychological Review, 1978, 85, 153-171.
- Frey, P. M., & Adelman, P. Recall memory for visually-presented chess positions. Memory and Cognition, 1976, 4, 541-547.
- Friedland, P. Knowledge-based hierarchical planning in molecular genetics. Ph.D. Thesis, Computer Science Department, Stanford University, 1979.

- Friedman, W. F., & Kirkpatrick, S. E. Congenital aortic stenosis. In A. J. Moss, F. H. Adams, & G. C. Emmanouilides (Eds.), Heart disease in infants, children, and adolescents. Baltimore, Md.: Williams and Wilkins, 1977.
- Glaser, R., & Pellegrino, J. W. Improving the skills of learning. Intelligence, in press.
- Goldin, S. E. Effects of orienting tasks on recognition of chess positions. American Journal of Psychology, 1978, 91, 659-672.
- Green, C. The application of theorem proving to question-answering systems. Ph.D. Thesis, Department of Electrical Engineering, Stanford University, 1969. Also, Stanford Artificial Intelligence Project Memo AI-96, June 1969.
- Greeno, J. G. Trends in the theory of knowledge for problem solving. In D. T. Tuma & F. Reif (Eds.), Problem solving and education: Issues in teaching and research. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1980.
- Hartmann, A. F., Goldring, D., Strauss, A. W., Hernandez, A., McKnight, R. C., & Weldon, C. S. Coarctation of the aorta. In A. J. Moss, F. H. Adams, & G. C. Emmanouilides (Eds.), Heart disease in infants, children, and adolescents. Baltimore, Md.: Williams and Wilkins Co., 1977.
- Hinsley, D. A., Hayes, J. R., & Simon, H. A. From words to equations: Meaning and representation in algebra word problems. In M. A. Just & P. A. Carpenter (Eds.), Cognitive processes in comprehension. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978.
- Johnson, P. E., Feltovich, P. J., Moller, J. H., & Swanson, D.B. Clinical expertise: Theory and data from the diagnosis of congenital heart disease. Paper presented at the meeting of the American Educational Research Association, 1979.(a)
- Johnson, P. E., Severance, D. G., & Feltovich, P. J. Design of decision support systems in medicine: Rationale and principles from the analysis of physician expertise. Proceedings of the Twelfth Hawaii International Conference on System Sciences, Hawaii, 1979. (b)

- Kassirer, J. P., & Gorry, G. A. Clinical problem solving: A behavioral analysis. Annals of Internal Medicine, 89, 245-255.
- Larkin, J. H. Skilled problem solving in physics: A hierarchical planning approach. Journal of Structural Learning, 1980.
- Lawlis, G. F., & Lu, E. Judgment of counseling process: Reliability, agreement, and error. Psychological Bulletin, 1972, 78, 17-20.
- Lesgold, A. M., Peltovich, P. J., Glaser, R., & Wang, Y. The acquisition of perpetual diagnostic skill in radiology. Tech. Report No. 1, ONR Contract No. N00014-79-C-0215. Pittsburgh, PA: LRDC, University of Pittsburgh, 1981.
- Lucas, R. V., & Schmidt, R. E. Anomalous venous connections, pulmonary and systemic. In A. J. Moss, F. H. Adams, & G. C. Emmanouilides (Eds.), Heart disease in infants, children, and adolescents. Baltimore, Md.: Williams and Wilkins, 1977.
- Martiniuk, R. G. Individual differences in motor performance and learning. In J. H. Wilmore (Ed.), Exercise and sport sciences review. N.Y.: Academic Press, 1974.
- McGuire, C., & Bashook, P. A conceptual framework for measuring clinical problem solving. Paper presented at the meeting of the American Educational Research Association, 1978.
- McGuire, C. H., & Solomon, L. Clinical simulation. New York: Appleton-Century-Crofts, 1971.
- Melton, A. W. Implications of short-term memory for a general theory of memory. Journal of Verbal Learning and Verbal Behavior, 1963, 2, 1-21.
- Minsky, M. A framework for representing knowledge. In P. Winston (Ed.), The psychology of computer vision. New York: McGraw Hill Inc., 1975.
- Moller, J. H. Essentials of pediatric cardiology. Philadelphia: F. A. Davis Company, 1978.
- Moller, J. H., Amplatz, K., & Edwards, J. E. Congenital heart disease. Kalamazoo, Mich.: The Upjohn Company, 1971.

- Moss, A. J., Adams, F. H., & Emmanouilides, G. C. Heart disease in infants, children and adolescents. Baltimore, Md.: Williams and Wilkins Company, 1977.
- Newell, A. Heuristic programming: Ill-structured problems. In J. Aronofsky (Ed.), Progress in operations research, vol. III. New York: Wiley, 1969.
- Newell, A. Artificial intelligence and the concept of mind. In R. C. Shank & K. M. Colby (Eds.), Computer models of language and thought. San Francisco: W. H. Freeman & Co., 1973.
- Newell, A., & Rosenbloom, Paul S. Mechanisms of skill acquisition, the law of practice, and production system architecture. In J. R. Anderson (Ed.), Cognitive Skills and their acquisition. Hillsdale, N.J.: Lawrence Erlbaum, in press.
- Newell, A., & Simon, H. Human problem solving. Englewood Cliffs, N.J.: Prentice-Hall, 1972.
- Norman, D. A., Rumelhart, D. E., & the LNR Research Group. Explorations in cognition. San Francisco: Freeman, 1975.
- Ornstein, P. A., & Corsale, K. Organizational factors in children's memory. In C. R. Puff (Ed.), Memory organization and structure. New York: Academic Press, in press.
- Paige, J. M., & Simon, H. A. Cognitive processes in solving algebra work problems. In B. Kleinmuntz (Ed.), Problem solving. New York: Wiley, 1966.
- Pellegrino, J. W., Chi, M. T. H., & Majetic, D. Ability differences and the processing of quantitative information. Paper presented at the 19th Annual Meeting of the Psychonomic Society, San Antonio, Texas, 1978.
- Piaget, J. The Psychology of Intelligence. Totowa, N. J.: Littlefield, Adams & Co., 1972.

- Polson, P., Atwood, M. E., Jeffries, R., & Turner, A. The processes involved in software design. In J.R. Anderson (Ed.), Cognitive skills and their acquisition. Hillsdale, N.J.: LEA, in press.
- Pople, H. E. The formation of composite hypothesis in diagnostic problem solving: An exercise in synthetic reasoning. Paper presented at the International Joint Conference on Artificial Intelligence, 1977.
- Quillan, R. The teachable language comprehender: A simulation program and theory of language. Comm. ACM, 1969, 12, 459-476.
- Reed, S. K. Category vs. item learning: Implications for categorization models. Memory and Cognition, 1978, 6, 612-621.
- Reitman, J. Skilled perception in GO: Deducing memory structures from inter-response times. Cognitive Psychology, 1976, 8, 336-356.
- Rosch, E., & Mervis, C.B. Family resemblance: Studies in the internal structure of categories. Cognitive Psychology, 1975, 7, 573-605.
- Rosch, E., Mervis, C. B., Gray, W., Johnson, D., & Boyes-Braem, P. Basic objects in natural categories. Cognitive Psychology, 1976, 8, 383-440.
- Rubin, A. D. Hypothesis formation and evaluation in medical diagnosis (MIT-AI Technical Report No. 316), Cambridge, Mass.: MIT, 1975.
- Rumelhart, D. E. Analogical processes and procedural representations. CHIP Report No. 81, Center for Human Information Processing, University of California, San Diego: February 1979.
- Rumelhart, D. E., & Norman, D. A. Accretion, tuning and restructuring: Three modes of learning. In R. Klatsky & J. W. Cotton (Eds.), Semantic Factors in cognition. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

- Rumelhart, D. E. & Ortony, A. The representation of knowledge in memory. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), Schooling and the acquisition of knowledge. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.
- Sacerdoti, E. D. A structure for plans and behavior. New York: Elsevier-North Holland Publishing, 1977.
- Schank, R. & Abelson, R. Scripts, plans, goals, and understanding. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.
- Shavelson, R. J. Some aspects of the correspondence between content structure and cognitive structure in physics instruction. Journal of Educational Psychology, 1972, 63, 225-234.
- Shortliffe, E. H. Computer-based medical consultations: MYCIN. New York: American Elsevier, 1976.
- Siegler, R. S. Three aspects of cognitive development. Cognitive Psychology, 1976, 8, 481-520.
- Siegler, R. S. The origins of scientific reasoning. In R. S. Siegler (Ed.), Children's thinking: What develops? Hillsdale, N. J.: Lawrence Erlbaum Associates, 1978.
- Simon, D. P., & Simon, H. A. Individual differences in solving physics problems. In R. Siegler (Ed.), Children's thinking: What develops? Hillsdale, N. J.: Lawrence Erlbaum Associates, 1978.
- Simon, H. A., & Chase, W. G. Skill in chess. American Scientist, 1973, 61, 394-403.
- Spilich, G. J., Vesonder, G. T., Chiesi, H. L., & Voss, J. F. Text processing of domain-related information for individuals with high and low domain knowledge. Journal of Verbal Learning and Verbal Behavior, 1979, 14, 506-522.
- Stevens, A., Collins, A., & Goldin, S. E. Misconceptions in student's understanding. International Journal Man-Machine Studies, 1979, 11, 145-156.

- Swanson, D. B. Computer simulation of expert problem solving in medical diagnosis. Ph.D. Thesis, University of Minnesota, 1978.
- Swanson, D. B., Feltovich, P. J., & Johnson, P. E. Are information processing theories relevant to instructional psychology? Paper presented at the meeting of the American Educational Research Association, 1977.
- Swanson, D., Feltovich, P. J., Johnson, P. E., & Moller, J. H. A computer simulation study of clinical expertise: Toward a knowledge-based definition of clinical competence. Paper presented at the meeting of the American Educational Research Association, 1979.
- Thorndyke, P. W., & Stasz, Cathleen. Individual differences in procedures for knowledge acquisition from maps. Cognitive Psychology, 1980, 12, 137-175.
- Thro, M. P. Relationships between associative and content structure of physical concepts. Journal of Educational Psychology, 1978, 70, 971-978.
- Tinsley, H. E. A., & Weiss, D. J. Interrater reliability and agreement of subjective judgments. Journal of Counseling Psychology, 1975, 22, 358-376.
- Tulving, E. Ecphoric processes in recall and recognition. In J. Brown (Ed.), Recall and recognition. London: Wiley, 1976.
- Tulving, E., & Pearlstone, E. Availability versus accessibility of information in memory for words. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 381-391.

- VanLehn, K., & Brown, J. S. Planning nets: A representation for formalizing analogies and semantic models of procedural skills. In R. E. Snow, P. A. Federico, & W. E. Montague (Eds.), Aptitude, learning and instruction: Cognitive process analyses. Hillsdale, N. J.: Lawrence Erlbaum Associates, 1979.
- Wason, P. C., & Johnson-Laird, P. N. Psychology of reasoning: Structure and content. Cambridge, MA: Harvard University Press, 1972.
- Wortman, P. M. Medical diagnosis: An information processing approach. Computers and Biomedical Research, 1972, 5, 315-328.
- Wortman, P. M., & Greenberg, L. D. Coding, recoding, and decoding of hierarchical information in long-term memory. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 234-243.

APPENDIX A

THE INFORMATION STRUCTURE OF A CASE DESCRIPTION

Age, weight, height, presenting problem

HISTORY

History of pregnancy, birth weight

Neonatal history

Childhood history (if applicable)

Family history

PHYSICAL EXAMINATION

General appearance

Fingers and toes

Respiration

Blood Pressure

Pulses

Chest examination

General appearance

Lungs

Palpation

First heart sounds, systolic murmurs

Second heart sounds, diastolic murmurs

Palpation of abdomen

X-RAY

Formal report

EKG

Formal report

APPENDIX B

Patient Files for the Five Cases of the Study

CASE 1

Discrete Membranous Subaortic Stenosis

DESCRIPTION

- (1) The patient is an 11 year old boy. He weighs 62 pounds and is 53 inches tall. His presenting problem is that a physician doing a school physical, three months ago, heard a murmur.

HISTORY

- (2) The child was born following a normal pregnancy and delivery. He was delivered at full term with a birthweight of 5 pounds, 6 ounces.
- (3) A murmur was heard at birth by the attending physician, but there has never been a follow-up of this murmur until the present.
- (4) Other than the murmur at birth, there were no problems in the neonatal period. Specifically, the child was not cyanotic in the neonatal period.
- (5) The boy led a normal, asymptomatic childhood until the age of 10 years when, during an examination for a school physical, a murmur was noted by the physician doing the exam and he referred the child to the University of Minnesota for evaluation.

- (6) His growth and development have been normal when compared to his siblings and he has done well in school.
- (7) There is no history of cyanosis or chest pains. He currently participates in baseball, football, and other sports and has no evidence of fatigue or dyspnea on exertion.
- (8) There have been no unusual childhood illnesses, and he has had no hospitalization or operations.
- (9) The family history is negative for congenital heart disease.

PHYSICAL EXAM

- (10) On physical examination, this is a healthy appearing, 11 year old boy. The facies are normal.
- (11) There is no cyanosis and the fingers and toes appear normal.
- (12) There is no evident respiratory distress. Respiration rate is 16/minute.
- (13) The blood pressure in the left arm is 90/70, in the right arm 102/60 and in the leg is 120/60 by auscultation.
- (14) The pulses are normal and equal in arms and legs. The pulse rate is 84/minute.
- (15) The chest is symmetrical with no bulge. There is mild asymmetry of the back associated with slight

thoracic scoliosis. The lungs are clear to auscultation.

- (16) The apex impulse is in the fifth intercostal space in the midclavicular line.
- (17) There is a systolic thrill felt below the right clavicle, along the mid-left sternal border, and in the suprasternal notch.
- (18) The first heart sound is normal and there is no systolic ejection click.
- (19) A grade 4 over 6 systolic ejection murmur is heard best at the right upper sternal border, but also heard along the left sternal border, and less well throughout the precordium. The murmur is also heard in the neck.
- (20) The second heart sound is normal and followed by a grade 1-2 over 6 early diastolic murmur heard best at the second left intercostal space.
- (21) Neither the liver nor spleen was palpable.

X-RAY

- (22) The chest x-ray shows normal cardiac size and contour and normal vascularity, but prominence of the ascending aorta.

AD-A105 753

PITTSBURGH UNIV PA LEARNING RESEARCH AND DEVELOPMEN--ETC F/G 6/5
KNOWLEDGE BASED COMPONENTS OF EXPERTISE IN MEDICAL DIAGNOSIS.(U)
SEP 81 P J FELTOVICH

N00014-79-C-0215

NL

UNCLASSIFIED

LRDC-81/PDS-2

3 OF 3

AD-A105 753



END

DATE

FILED

11 81

DTIC

EKG

- (23) The EKG shows a QRS axis of between $+30^{\circ}$ and $+45^{\circ}$. The P waves, T waves, and PR interval are normal. The pattern shows probably left ventricular hypertrophy.

CASE 2

Total Anomalous Pulmonary Venous Connection

DESCRIPTION

- (1) The patient is a 5 year old girl. She weighs 33 pounds and is 41 inches tall. Her presenting problem is a murmur heard by a family physician.

HISTORY

- (2) The child was born following a normal pregnancy, labor, and delivery. She was delivered at full term with a birthweight of 7 pounds, 4 ounces.
- (3) A murmur was heard in the first day of life by the attending physician. He continued to see the child periodically and referred the child to the University of Minnesota hospitals at age 5 years.
- (4) Other than the murmur, there were no problems in the neonatal period. Specifically, the child was not cyanotic in the neonatal period.
- (5) Infancy was unremarkable. However, between the ages of 2 and 3 years the patient had numerous infections including flu, upper respiratory infections, and otitis, and required several hospitalizations for respiratory infections in her home town.

- (6) In the past two years, her health has been good. There have been no complaints of chest pain, dyspnea or palpations. She has always preferred quiet activities and declines invitations by peers to participate in active sports.
- (7) The mother, who has two younger children, has noted that in the last two years, when the child is cold, her lips turn blue. In the last year, the mother has noted dusky nail beds but cannot relate this to any specific conditions.
- (8) Development has been normal throughout life but growth has been slow.
- (9) Aside from a paternal aunt who has an asymptomatic heart murmur, there is no family history of heart disease. The child's two younger siblings are asymptomatic.

PHYSICAL EXAM

- (10) On physical examination, this is an "anxious" appearing 5 year old. The facies are normal.
- (11) She shows minimal circumoral cyanosis. Fingernails appear minimally cyanotic with slight watch-crystal formation of the fingernails.
- (12) There is no evident respiratory distress. Respiration rate is 20/minute.

- (13) The blood pressure is 106/60 in the right arm, 108/64 in the left arm and 114/72 in the leg by auscultatory method.
- (14) The pulses are full and equal in arms and legs. The pulse rate is 110/minute.
- (15) Examination of the chest shows a prominent precordial bulge. The lungs are clear to auscultation.
- (16) The apex impulse is felt lateral to the mid-clavicular line in the 6th intercostal space.
- (17) Auscultation of the heart shows a first heart sound with a very loud component. The second heard sound is widely split all the time and appears fixed. The pulmonary component is a little prominent.
- (18) A grade 2-3 over 6 systolic ejection type murmur is present along the upper left sternal border, the murmur being as loud or perhaps even louder over the left upper back.
- (19) A grade 2 over 6 mid to late diastolic murmur is present along the left sternal border.
- (20) The liver and spleen are not palpable.

X-RAY

- (21) The chest X-ray shows moderate cardiomegaly and markedly increased pulmonary vasculature. There is insufficient amount of barium in the esophagus on the lateral view to adequately evaluate the size of the left atrium. There is an unusual shadow seen in the right side representing, very likely, an anomalously coursing pulmonary vein.

EKG

- (22) The EKG shows right axis deviation of $+120^{\circ}$, a wandering atrial pacemaker, right atrial enlargement, and right ventricular hypertrophy. There is an rsR's' pattern in lead V_1 .

CASE 3

Patent Ductus Arteriosus

DESCRIPTION

- (1) The patient is a 6 month old girl. She weighs 18 pounds, 10 ounces, and her length is 27½ inches. Her presenting problem is a murmur.

HISTORY

- (2) The child was born following a pregnancy in which there was "flu" at 2½ months of gestation. Otherwise, labor and delivery were unremarkable. She was delivered at full term with a birth weight of 7 pounds 8 ounces.
- (3) A cardiac murmur was first heard at the age of 2½ weeks by her physician in Colorado Springs, Colorado. Otherwise, there were no problems in the neonatal period.
- (4) The family then moved to Minnesota and the murmur was again heard at age 4 months by their pediatrician who referred her to the University of Minnesota hospitals.
- (5) There have been a paucity of cardiac symptoms. The mother, who has one older child, has noted no real differences between this child and her other child.

- (6) There have been no difficulties with feeding, that is, no excessive perspiration or slowness.
- (7) On questioning, mother says occasionally the baby is blue around the mouth during exertion.
- (8) There have been no unusual number of respiratory infections.
- (9) The family history is negative for congenital heart disease.

PHYSICAL EXAM

- (10) On physical examination, this is a "chubby", healthy looking baby. The facies are somewhat flat and show broad nasal bridge but are otherwise normal.
- (11) There is no cyanosis and the fingers and toes are normal.
- (12) There is no evident respiratory distress. The respiratory rate varies from 40 to 44/minute.
- (13) The simultaneous flush blood pressures in the right arm and leg are 90.
- (14) The peripheral pulses seem brisk in both arms and the legs. The pulse rate varies from 120 to 160/minute.
- (15) The chest is symmetrical. The lungs are clear to auscultation.

- (16) The apex impulse is located in the fourth intercostal space in the mid-clavicular line.
- (17) There are no heaves or thrills.
- (18) The first and second heart sounds are normal.
- (19) A grade 2 over 6 systolic murmur coupled to a grade 1 over 6 diastolic murmur is heard in the left infraclavicular area. The murmur has the characteristics of a machinery murmur. It reaches its peak intensity at about the second heart sound.
- (20) The liver is palpable 1 cm. below the right costal margin and the spleen tip at the left costal margin.

X-RAY

- (21) The X-ray shows normal cardiac size and contour. The pulmonary vasculature is normal or slightly increased.

EKG

- (22) The EKG has voltage criteria for left ventricular hypertrophy.

CASE 4

Pulmonary Atresia

DESCRIPTION

- (1) The patient is a 4 day old male. His weight is 7 pounds 4 ounces and his length is 20 inches. His presenting problem is cyanosis.

HISTORY

- (2) The child was born at term following a normal pregnancy and delivery. The mother had no unusual health problems during pregnancy.
- (3) The child was cyanotic at birth and because of cyanosis which persisted in oxygen and also mild respiratory distress, he was transferred to St. Paul Children's Hospital.
- (4) On the third day of life, a murmur was initially heard and because of this he was transferred to the University of Minnesota Hospitals.
- (5) Feeding was started on the first day of life and went well although the child regurgitated on several occasions.
- (6) The family history is negative for congenital heart disease.

PHYSICAL EXAM

- (7) On physical examination, the child is vigorous

and has a good cry. The facies are normal.

- (8) He is markedly cyanotic even in an atmosphere of 100% oxygen. The cyanosis is generalized.
- (9) Respiratory rate varies from 50 to 100/minute.
- (10) The blood pressures in the arm and leg are 40 millimeters mercury by simultaneous flush methods.
- (11) The peripheral pulses are normal and equal in arms and legs. The pulse rate is 130/minute.
- (12) The chest is symmetrical. On auscultation, the lungs show no rales or rhonchi.
- (13) There are no heaves or thrills. The apex impulse is in the 4th intercostal space in the mid-clavicular line.
- (14) The first heart sound is loud and there is a grade 2 over 6 systolic murmur along the lower left sternal border. There is little radiation. This murmur is not heard in the back or head.
- (15) The second heart sound is single and perhaps slightly increased in intensity. There is no gallop or diastolic murmur.
- (16) The liver edge is palpable 2 cm. below the right costal margin. The spleen is not palpable.

X-RAY

- (17) The X-ray shows the heart to be somewhat enlarged with the right atrium also noted to be enlarged.

The pulmonary vasculature is diminished. There is no evidence of pulmonary infiltrate. The base of the heart is narrow.

EKG

- (18) The EKG has a QRS axis of $+50^{\circ}$ and shows sinus tachycardia, right atrial enlargement, and non-specific ST and T wave changes. In addition, the EKG shows a small R wave in the right precordium suggesting or compatible with a hypoplastic right ventricle.

CASE 5

Coarctation of the Aorta and Bicuspid Aortic Valve and Mitral Insufficiency

DESCRIPTION

- (1) The patient is a 9 month old boy. He weighs 18 pounds, 2½ ounces and is 28½ inches long. The presenting problem is a heart murmur noted by a private physician three weeks ago on a routine examination.

HISTORY

- (2) The child was born following a normal pregnancy and delivery. He was delivered, along with a twin, at full term with a birth weight of 6 pounds 14 ounces. The twin is asymptomatic. There is a healthy 4 year old brother.
- (3) There were no problems at birth or in the neonatal period. Specifically, there were no murmurs heard or cyanosis in the neonatal period.
- (4) Although the child had been checked periodically, no murmur was noted until three weeks ago. Cyanosis has never been noted.
- (5) There have been no problems with feedings, that is, no unusual perspiring, tiring, or slowness. His suck is good.

- (6) There have been no unusual illnesses, in particular, no recurrent respiratory infections.
- (7) His growth and general development have been normal when compared to his 4 year old brother. He is active but mother thinks somewhat less active than his twin or 4 year old brother.
- (8) The family history is negative for congenital heart disease.

PHYSICAL EXAM

- (9) On physical examination, this is a well-developed, well nourished alert 9 month old child. Facies are normal.
- (10) There is no cyanosis and fingers and toes are normal.
- (11) There is no evident respiratory distress. Respiratory rate is 16/minute.
- (12) Blood pressures were 140 in the left arm and 80 in the left leg by simultaneous flush method. Blood pressure in the right arm by auscultation 140/90. Blood pressure in left arm by auscultation 160/90.
- (13) The radial pulses are bilaterally sharp and readily palpable. Femoral pulses are not palpable bilaterally. Pulse rate is 110/minute.

- (14) The neck is supple with no masses. There is no thrill in the suprasternal notch.
- (15) The chest is symmetrical. The lungs are clear to auscultation.
- (16) The apex impulse is in the fifth intercostal space in the mid clavicular line.
- (17) There are no heaves or thrills.
- (18) The first heart sound is split with the second component perhaps representing an apical systolic ejection click. The second heart sound is split and P_2 is slightly accentuated.
- (19) There is a grade 3 over 6 soft systolic murmur at the apex which radiates to the axilla.
- (20) There is a grade 1 to 2 over 6 short, harsh systolic murmur heard best in the aortic area. This murmur is not well heard in the interscapular area.
- (21) The liver is palpable 2 cm below the right costal margin and the spleen is not palpable.

X-RAY

- (22) The X-ray shows moderately enlarged heart and left atrial enlargement. The pulmonary vasculature is prominent and perhaps slightly increased.

EKG

(23) The EKG shows biventricular hypertrophy and left atrial enlargement. The axis is normal.

END

APPENDIX C

INSTRUCTIONS TO SUBJECTS

This is a study of diagnostic thinking. I will give you exercises in which you are to reach a diagnostic conclusion given a set of data from a patient case.

The data I will give you are based upon actual cases. They consist of written statements, developed by a faculty consultant to the project, summarizing each of the major categories of patient findings. History statements were developed by the consultant based on his reading of the patient chart. Physical examination statements represent conclusions reached by a pediatric cardiologist who originally conducted the examination. X-ray and EKG statements are from the "formal reports" which are part of the patient file.

The exercises I will present all represent instances of congenital heart disease or conditions often confused with congenital heart diseases. Therefore, some data that you would normally find in a general pediatric work-up have been omitted. All screening of data was done by the faculty consultant. If some datum of interest to you is not eventually reported, assume that it was found to be normal. The

diseases in the study may differ in how common they are, how complex they are, etc. Hence, you should try to diagnose each case independently of the others.

For each exercise, I will give you a "patient file" consisting of typed sheets of patient data. The data are typed in the order: general patient description, history, physical examination, X-ray and EKG. Within these major categories, data are segmented into small numbered groups.

When I give you the patient file, I would like you to open it and read out loud each numbered data group starting with the first. For each numbered data group, read its number followed by the data itself. Then indicate when you have finished reading the group by saying "period" or "stop."

After you read a numbered data group, please think out loud about its significance toward formulating a diagnosis for the patient. When you have finished thinking about a data group, go on and read the next.

Please try to be as thorough as possible in reporting your thoughts as they arise, even if they seem unimportant to you. In particular, try to make clear when you first think of something, for example, a possible diagnosis, whether data are consistent or

inconsistent with "hunches" you have, and when you eliminate a diagnosis you had been considering.

As you read the patient file please do not "skip ahead"; it is important that you consider the data in the order in which they appear.

If, at any time, you want to review any data group you may do so. If you return to previously considered data, please read it aloud again so that I will know what data you are considering. After you have re-read and thought about data please return to the next new data group.

At four points during each exercise, after history, physical, X-ray, and EKG I will say "Please tell me about hunches." At these points, I would like you to just tell me what diagnoses (if any) you are actively considering for the patient at the time I interrupt. At these times I do not want you to do a great deal of additional thinking beyond what you have already done when I interrupt. The purpose of my "probe" is simply to get an explicit listing of the hypotheses you are considering. Report your hypotheses in the manner that best represents the way you are thinking about them. If you have no hypotheses or "hunches" when I interrupt, say so and go on.

Throughout the exercise, whenever I judge that an unusual amount of time has passed without your saying anything, I will say "Please talk more." This is just to encourage you to report your thoughts.

At the end of each exercise, I will ask you to give a primary diagnosis. This is the diagnosis you think is the best description of the patient's condition. I will also ask you to give secondary diagnoses. These are diagnoses you feel might apply to the patient, but about which you are not as confident as you are about the primary diagnosis. You may give as many as two secondary diagnoses; you may also give one or none.

I will give you the "catheterization diagnosis" for each case after you have completed all the cases of the study.

This is a research project and not a test. Your participation will be confidential as described in the consent form; hence, I hope you will be relaxed in doing the exercise.

Do you have any questions?

APPENDIX D
INTERRATER AGREEMENT

A Study of Interrater Agreement in Coding All
Hypotheses from the Protocols

A study was completed to determine the degree to which three independent raters agreed in their interpretations and subsequent scoring of statements generated by subjects while diagnosing cases of congenital heart disease. Subjects were presented with medical case data in the area of pediatric cardiology and were asked to speak aloud as they considered the given information while making a diagnosis. The resulting protocols of subjects' performance (e.g., Appendix H), contain the hypotheses, comments, and diagnoses from twelve subjects: four experts, four residents ("trainees") in the area of pediatric cardiology, and four medical students. If high or moderately high interrater agreement could be found for a chosen subset of transcripts, then the investigator could state with some degree of confidence that the scoring of the protocols is consistent across transcripts and between raters.

Tinsley and Weiss (1975), proposed that interrater agreement reflects the extent to which different judges tend to make exactly the same judgments about a given subject response. Interrater agreement rather than

interrater reliability was selected as the appropriate index, as interrater reliability measures the degree to which the ratings of different judges are proportional when expressed as deviations from their means (Tinsley and Weiss, 1975). In this study, the investigator was interested in the absolute agreement between raters rather than the amount of deviation from a mean. The interrater agreement coefficient considers the percentage of agreement between raters and adjusts this percentage by removing the percentage of agreement predicted by chance. This coefficient, Cohen's k, is represented as:

$$k = \frac{P_o - P_c}{1 - P_c}$$

where,

P_o = percentage or proportion of ratings in which the two judges agree

P_c = percentage or proportion of ratings for which agreement is expected by chance

The figures used for chance probability of interrater agreement were prepared by Tinsley and Weiss based on formulas appearing in an article by Lawlis and Lu (1972). The chance probability figures used were based on a zero point discrepancy between judges.

Six cases (one, VCLAC, was not included in the present report), were presented to each of the

twelve subjects previously described. From the seventy-two resulting protocols, thirteen (18%) were selected for inclusion in the agreement study. The reports were chosen so that each subject and case was examined at least once. An effort was made to include an equal number of experts, midrange ("trainees") and novices (medical students) in the study. Tables 1-D and 2-D indicate the nature of the cases and subjects selected.

Table 1-D
Frequency of Subjects Examined
Within Each Case

<u>Case</u>	<u>Frequency of Subjects Examined</u>
TAPVC	3 (2 experts, 1 novice)
Pulmonary Atresia	2 (1 expert, 1 novice)
VCLAC	2 (1 midrange, 1 novice)
Coarctation	2 (2 midrange)
Subaortic Stenosis	1 (1 expert)
Patent Ductus	3 (1 expert, 1 midrange, 1 novice)

Table 2-D
Distribution of Subject Level
of Expertise

<u>Subject Level of Expertise</u>	<u>Frequency</u>
Expert	5
Midrange	4
Novice	4

Independent ratings were made by two of the three judges for each of the thirteen reports selected. Each judge listed in order all of the hypotheses used by subjects that she identified in the examination of the protocol. A comparison was made of these hypothesis lists and the proportion of agreement between raters was determined (Table 3-D). The total number of hypotheses for each subject indicates a combined set of common and unique hypotheses generated by both judges.

The interrater agreement index, Cohen's K, was employed to adjust for agreement due to chance. The results are shown in Table 4-D. The results indicate a high degree of interrater agreement between independent judges of the hypotheses generated by the subjects in the protocol. A visual comparison of the proportion of agreement and interrater agreement coefficients suggests that the probability of agreement by chance is small. This is indicated by only a slight decrease in the interrater agreement coefficients (a mean decrease of .01).

Although the numbers of subjects within each level of medical skill is small, the averages in Tables 5-D and 6-D suggest that the interrater agreements within skill level and within cases is high.

Table 3-D
Proportion Agreement Between Raters on Cases

Case	Subjects	Proportion Agreement
TAPVC	E1	38 of 46 = .83
	S4	21 of 28 = .75
	E2	18 of 25 = .72
Pulmonary Atresia	E4	17 of 21 = .81
	S2	24 of 27 = .89
VCLAC	T3	15 of 24 = .63
	S1	9 of 15 = .60
Coarctation	T4	21 of 27 = .78
	T1	10 of 14 = .71
Subaortic Stenosis	E3	20 of 24 = .83
Patent Ductus Arteriosus	T2	8 of 11 = .73
	E4	7 of 8 = .88
	S3	8 of 9 = .89

Table 4-D
Interrater Agreement Coefficients for Cases

Cases	Subjects	Cohen's K
TAPVC	E1	K = .83
	S4	K = .74
	E2	K = .71
Pulmonary Atresia	E4	K = .80
	S2	K = .89
VCLAC	T3	K = .61
	S1	K = .57
Coarctation	T4	K = .77
	T1	K = .69
Subaortic Stenosis	E3	K = .82
Patent Ductus Arteriosus	T2	K = .70
	E4	K = .86
	S3	K = .88

Table 5-D

Mean Interrater Agreement
Coefficient with Skill Level

<u>Skill Level</u>	<u>N</u>	<u>\bar{X} Coefficient</u>
Expert	5	.80
Midrange	4	.69
Novice	4	.77

Table 6-D

Mean Interrater Agreement
Coefficient with Case

<u>Case</u>	<u>N</u>	<u>\bar{X} Coefficient</u>
TAPVC	3	.76
Pulmonary Atresia	2	.85
VCLAC	2	.59
Coarctation	2	.73
Subaortic Stenosis	1	.82
Patent Ductus	3	.81

To summarize, a study was completed to compare the degree of interrater agreement between two independent judges for a set of introspective reports. After analyzing the scoring of the different judges, it was concluded that a high level of interrater agreement exists both across subject level of skill and within each case.

APPENDIX E

SCORING INSTRUCTIONS

CASE 1

(General)

All scoring that you do should be as "blind" as possible as to the identity and level of experience of the subject being scored. Hence, you should assemble all subjects' protocols for the case and ask some other person to cover up the identifying markings that appear on each page of each protocol (e.g., S7) so that these markings are not visible when you score. A suggestion is to use a small piece of paper and some tape so that these covers can be removed later without damage to the protocol. You should also shuffle the pile of protocols before scoring so that there will be no systematic order.

A set of scoring instructions will be provided for the case. You should first read these instructions to get a general idea of what you are to be doing. Then, it is equally important that you read each protocol once or twice before you do any scoring to get a general "feel" for what the subject is doing on the case.

Practice trials and procedure

For all scoring to be done on the case, there will be a small number (two subjects) of practice scoring trials that will give us an opportunity to compare our scoring before you score the remainder of cases by yourself. The procedure to be followed on these practice trials is as follows:

- 1) Read scoring instructions for each type of scoring you are to do on the case.
- 2) Score for two subjects whom I will suggest. Practice subjects are chosen because they provide good examples of the scoring decisions to be made. For practice scoring trials it will be necessary for you to know the number (e.g., S7) of the subject you are scoring.
- 3) After scoring, compare your scoring on practice subjects to mine (which will be provided) and read my comments on my scoring decisions (also provided) for the practice subjects. Make notes on discrepancies in scoring.
- 4) When you have done the above for the practice subjects, call me so that we can

discuss any problems you had in scoring and discrepancies in scoring.

Production Scoring

After the practice period for the case, you will score all subjects (including trial subjects) blind and using only the instructions and materials provided. On the basis of practice scoring trials, some modifications of scoring instructions may be made.

APPENDIX F

SCORING INSTRUCTIONS

Subvalvular Aortic Stenosis: Use of Aortic Stenosis Variants

There are two objectives of the scoring associated with this table. The first is to document the variants of Aortic Stenosis that are used together by the subject at the time of presentation of a numbered data item and at the times when the subject reports hunches, and to document the first mentioning (triggering) of each variant by the subject. The second is to document the evaluative "tone" with which the subject mentions the variant.

As with other scoring on the Aortic Stenosis case, the following correspondences hold:

Valvular Aortic Stenosis (ValvAS) = "aortic stenosis" when it is given without qualification as to type (Valvular AS is taken to be the default value of "aortic stenosis" when this term is left unqualified), "valvular aortic stenosis," "aortic valvular stenosis," "bicuspid aortic valve," and other statements that clearly place a stenosis at the level of the aortic valve.

Subvalvular Aortic Stenosis (SubAS) = "sub-valvular aortic stenosis," "membranous aortic stenosis," "membranous stenosis," and other statements that clearly place a stenosis below the level of the aortic valve.

Supravalvular Aortic Stenosis (SupAS) = "supra-valvular aortic stenosis" and other statements that clearly place a stenosis above the level of the aortic valve. These are not to be confused with "coarctation of the aorta" and "interrupted aortic arch" which are different diseases.

Scoring is directed at whether or not each of the three variants of Aortic Stenosis is mentioned by a subject at the time of presentation of each numbered data item (and at four points of solicitation of hunches) and, for each mentioned, whether it is mentioned with a positive (+), negative (-), or neutral (o) valence as a possible explanation for the patient's condition. The first mention (triggering) of a variant is a mention like others and is to be so coded. However, a special symbol (§T) will be used for the first mention of a variant to signify its special status. Triggerings will also be coded for evaluative valence (+, -, o).

Scoring Procedure

A table (Table 1) is provided which has three coding lines for each subject with each line labelled at the left with either ValvAS, SubAS, or SupAS. Across the top are codes for the twenty-two data items and the four places of solicitation of hunches, all arranged in order from left to right.

The unit of analysis for the scoring is the subject's entire response after his reading of a numbered data item. Read each numbered data item (and "hunches" at the appropriate points) starting from the beginning of the case. After reading each data item, then read the subject's entire response following the data item and decide which, if any, of the variants of Aortic Stenosis have been mentioned by the subject. Then, for each variant mentioned, decide if it is mentioned with a positive, negative, or neutral evaluative valence as a possible explanation for the patient's condition and place a corresponding +, -, o, on the appropriate line for the variant and under the column heading for the data item as shown below.

		(1)	(2)
	ValvAS	+	
S1	SubAS	+	-
	SupAS		-

If the same variant is mentioned more than once in the same statement, the "net" valence must be determined (see guidelines below).

The first mention (triggering) in the case of each variant is to be marked with a "T" rather than a simple +, -, o. As with other mentions, you should decide whether the variant is being raised with a positive, negative, or neutral evaluative valence and place a corresponding symbol (+, -, o) to the right of the "T" (for example "T+"). The first mention of a variant can occur at "hunches" as well as at the time of reading a numbered data item.

Keep in mind that the objective of Table 1 coding is to record those variants of AS that are mentioned together at the time of presentation of a data item. For purposes of this table, it does not matter, for instance, that in mentioning an AS variant at, say 18, the subject is functionally responding to (e.g., evaluating, triggering with respect to) some other data item, say, 17. The straightforward rule for this table is: "If he says it at 18, code it at 18."

Guidelines and Cautions

I. "Mention" just means mention

This scoring should be relatively easy since we are just looking for the variants of AS that are mentioned in any way in the subject protocols that follow each data item.

II. Judging +, -, 0 in recording the variants mentioned

When a variant of AS is mentioned, it can be used as a possible explanation for the case with an evaluative valence that runs from extreme negative (e.g., "That suggests we are surely not dealing with SubAS") to extreme positive ("Until proved otherwise, now, the child must have valvular aortic stenosis.") Many statements fall somewhere in between. When you score a statement, make every effort to decide if the statement has a basically positive or negative evaluative loading. The zero (or neutral) convention should be used sparingly and reserved for those instances where you simply cannot make a positive-negative discrimination. Two kinds of statements seem most suited to the zero convention and examples of each are given below:

Puzzled thoughts and slips of the tongue

- (10) On physical exam, this is a healthy appearing boy. The facies are normal.

S. Ok, I think, ah, this is sort of important again in ruling out certain congenital problems. I think maybe more helpful, is the fact that the facies are normal, ah. In the category of diseases I'm looking at, I believe it is one form of aortic stenosis, I don't remember if it is supra or subvalvular, but ah, there is a specific facies to ah, that form of aortic stenosis, and I think this would be a consideration so that this would be helpful in ruling out that particular form. I believe, I think it is supravulvular.

Subvalvular Aortic Stenosis is mentioned here for the first time and must be coded with a T. The evaluative valence, however, seems impossible to determine and the zero convention should be used. Supravulvular is also mentioned here for the first time, and the valence seems to be negative. Although the words "aortic stenosis" appear, it seems clear that the only function of these words is in conjunction with "supra" or subvalvular"; hence, Valvular Aortic Stenosis is not coded separately.

Double evaluations, within the same statement, that seem to cancel each other out

(17) There is a systolic thrill felt below the right clavicle, along the mid-left sternal border, and in the suprasternal notch.

S. Uh, the presence of a thrill in those areas really suggests that this child has aortic stenosis. Uh, and (I can't tell from that description whether it is subvalvular, supra-valvular, or valvular aortic stenosis), (although the murmur in subvalvular aortic stenosis is usually located more towards the mid-left sternal border than the upper right sternal border).

Subvalvular AS is mentioned here for the first time. The statement in the first parentheses seems to have positive valence for SubAS since the subject says the thrill cannot be used to discriminate among the three variants. On the other hand, the second statement (second parentheses) seems somewhat negative for SubAS since the thrill is not exactly where he would like to see it. Again, this is an appropriate situation for

the zero convention on SubAS, SupAS and ValvAS are also mentioned here.

III. Variants mentioned at "hunches" points

Variants that are used at the times when the subject offers hunches have a special evaluative status. The fact that they are raised as "hunches" at all would seem to land some positive valence to their evaluative dimension - almost no matter how they are discussed as hunches. Hence, if a subject says "x, y, and z are possible and q is extremely unlikely", although the wording itself for "q" seems negative, the fact that "q" is being raised at all at the point where the subject's job is to report hunches overrides the wording itself. Short of absolutely negative statements (such as: "It surely cannot be SubAS"), almost all uses that occur at "hunches points" should be assigned a positive valence.

IV. Levels of ambiguity in referencing AS variants

LEVEL 1 (Little Ambiguity)

These are fairly direct statements of the AS variant (e.g., "aortic stenosis", "SubAS", etc.) and should cause little problem. The conventions for correspondence, given at the beginning of this set of

rules, should be followed. Remember, "aortic stenosis" without further qualification, is "ValvAS" in the scoring.

LEVEL 2 (Stenosis mentioned with location discernible):

At times, "stenosis" (or perhaps some variation like "obstruction") is discussed, the location of the stenosis is not directly given, but the statement, as a whole, can be used to discern the location.

Eg. 1 (23) The EKG shows a QRS axis of between $+30^{\circ}$ and $+45^{\circ}$. The P waves, T waves, and PR interval are normal. The pattern shows probable left ventricular hypertrophy.

S: The EKG is compatible with the presence of a stenotic or an obstruction at the aortic valve level. It is not diagnostic by any means. But, it helps to support the previous concepts of a stenosis at the aortic level, at aortic valve level, or immediately above it

This statement would be scored for both Valvular AS and Supravalvular AS (from the phrase "immediately above it").

Eg. 2 (22) The chest x-ray shows normal cardiac size and contour and normal vascularity, but prominence of the ascending aorta.

S: OK, that all fits in with, this is probably post-stenotic aortic dilatation, which means uh, well, he is twelve years old, so (it) is probably moderately stenosed.

The problem is to determine what the referent of "it" is. The related discussion of "post-stenotic aortic dilatation" is enough to localize the referant to some kind of Aortic Stenosis, but is not sufficient to localize beyond that (that is, not to Sub or Supra Valvular), Hence, by the default conventions, the statement should coded as ValvAS.

LEVEL 3 (Referents outside the statement):

Some subject statements make indirect reference to something outside of the statement being scored (e.g., "That goes along with what I said before"). It has been decided not to automatically exclude all such statements as too ambiguous to code. The guideline is to assign the referent only if you are quite certain as to its identity. Err on the side of not coding; if there is any question in your mind, do not code.

A rule of thumb might be to assign the referent if you are sure "beyond a reasonable doubt."

In attempting to assign such a referent, it is valid to use the case context surrounding the statement which contains the "outside" reference; however, in doing so, do not consult farther than one statement back and one statement forward.

The following is an example of an outside reference that seems clear enough to code. The outside reference is in the second statement, 18.

{17) There is a systolic thrill felt below the right clavicle, along the mid-left sternal border, and in the suprasternal notch.

S: Ok, I would ah, I guess this would be ah, somewhat suspicious of aortic stenosis. The fact that there is a systolic thrill below the right, yeah, below the right clavicle, Ok, that is ah, in the aortic area where I would suspect, where I would expect to find it if this were aortic stenosis. So I think this would be a very significant finding, in helping me to zero in a little bit stronger on aortic stenosis as my prime possibility.

(18) The first heart sound is normal and there is no systolic ejection click.

S: Ah, the first heart sound is normal. That, I would expect, ah. The fact that there is no systolic ejection, ah, would ah. That would go along too. I'm not really expecting ah, a volume overload, so I guess that would again go along with what I'm thinking.

(19) A grade 4 over 6 systolic ejection murmur is heard best at the right upper sternal border but also heard along the left sternal border and less well throughout the precordium. The murmur is also heard in the neck.

S: OK, I think this is a very good description of the murmur of aortic stenosis and this would be very helpful information for me to ah, pointing in on aortic stenosis.

The problem referent is the one referred to by the phrase "that would go along with what I'm thinking," which appears at 18.

At 17, the subject has stated that 17 was a "very significant finding" that helped him set up Aortic Stenosis as his "prime possibility". At 19, Aortic Stenosis is, again, the only disease candidate that appears. Sandwiched, as it is, between two strong

assertions for Aortic Stenosis there seems little doubt that the referent at 18 is the same.

LEVEL 4 (Things too nebulous to code):

Using the "beyond a reasonable doubt" criterion, there will be statements judged too ambiguous to be scored as ValvAS, SubAS, or SupAS. This is a judgement call and we must rely on the judgement of the scorer. The following seems to be an example of such a statement:

(17) There is a systolic thrill felt below the right clavicle, along the mid-left sternal border, and in the suprasternal notch.

S: Systolic below the right clavicle (pause) hmm, in the suprasternal notch, that (pause) a systolic thrill below the right clavicle. Well, when you have a ductus you can feel a thrill below the left clavicle, or a coarc. Maybe this indicates that he has a right aortic arch and ah, a coarc or a coarc of his right arch or some other aortic outflow obstruction. A suprasternal notch thrill, again, is well known to accompany ah, left ventricular outflow obstruction.

"Aortic outflow obstruction" and "left ventricular outflow obstruction" are too ambiguous, here, to code as any variant of Aortic Stenosis since the subject could conceivably mean other things (e.g., various kinds of interruptions of the aortic arch).

V. Inferences from negations

No inference should be made from the negation of one AS variant to the assertion of another; that is, "not valvular aortic stenosis" should not, in itself, be taken as evidence for the mentioning, by the subject, of some other variant.

VI. The "non-valvular" convention

When a subject refers to "non-valvular aortic stenosis", an attempt should be made to determine which of the other aortic stenosis variants he means. When you cannot make this decision with good confidence, leave the subject statement unscored.

VII . Concatenations of any Aortic Stenosis variant with Aortic Insufficiency (AI)

Whenever a subject conjoins Aortic Insufficiency with any one of the Aortic Stenosis variants (a conjoint hypothesis, ASAI), it is only the Aortic Stenosis part of the hypothesis that is the target of scoring in Table 1. Use the same rules for scoring this part of

the conjoint hypothesis that are to be used for scoring
Aortic Stenosis variants when they occur by themselves.

APPENDIX G

CASE 1

Coding by the Second Scorer for Table 1

Table 1-G gives the final coding for Table 1 in the body of the report by a second scorer other than the author. The two scorers, the author and the alternate, recorded a combined total of 168 uses of the three Aortic Stenosis variants across subjects. With regard to whether each variant was mentioned at the time of a particular data point, the two scorers agreed on 161 (96%) of the codings. The cells with a discrepancy are marked with a "check" in Table 1-G. A check indicates that the author judged a variant to be used in that cell while the other scorer did not, or vice-versa. None of these discrepancies would affect any analysis or interpretation based on Table 1 in the body of the report.

Table 1-G
Case 1: Subject Use of LCS Hypotheses in Response to Patient
Data Items as Judged by Alternate Scorer

Patient Data Items																						
Subject/Hypotheses	History					Physical Exam										X-ray		EKG				
	1	3	4	7	8	HHx	10	13	14	17	18	19	20	HPEx	22	HXray	23	HEKG				
S1 ValvAS			X	J		X		X			X	X		X	X	X	X	X				
SupAS							X															
S2 ValvAS				X		X	X			X	J	X	X	X	X	X	X	X				
SubAS							X															
SupAS							X															
S3 ValvAS					X	X		X	X	X	X	X	X	X	X	X	X	X				
S4 ValvAS												X		X	X	X	X	X				
T1 ValvAS					X				X	X	X	X	X	X	X			X				
SubAS												X	X	X	X	X	X	X				
SupAS																		X				
T2 ValvAS												X	X	X	X	X	J	X				
T3 ValvAS				X	X	X	J		X	X	X	X	X	X	X	X	X	X				
SubAS				X		X				X			X	X	X	X	X	X				
SupAS										X												
T4 ValvAS									X	X	X	X		X	J	X	X	X				
E1 ValvAS	X	X		X		X				X	X	X	X	X	X	X	X	X				
SubAS										X	X	X	X	X	X	X						
SupAS										X		X	X	X	X							
E2 ValvAS								X			X	X	X	X	J	X	X	X				
SupAS												X	X	X								
E3 ValvAS		X				X				X	X	X	X	X	X	X	X	X				
SubAS										X	X	X	X	X	X	X		X				
SupAS										X		X	X	X	X	X						
E4 ValvAS										X	X	X	X	X	X	X	X	X				
SubAS											X	X	X	X	X	X		X				
SupAS												X	X	X	X	X		X				

Note: (X) indicates a subject's use of a hypothesis at the time of a patient data item.
(J) indicates a point of discrepancy between the two scorers.
• The two experts with more than twenty years of experience.

APPENDIX H

SCORING DISCUSSION

Case 1 - Table 1

Subject T3

This Appendix gives the complete protocol of subject T3 from the diagnosis of Case 1. After each numbered data item from the Case, the subject's response is indicated by "T3". The scoring decisions of the author in relation to Table 1 in the report body, are indicated after the subject's response ("D1" etc.), along with the discussions of these decisions which were presented to the second scorer as part of the training procedure for scoring (see Appendix E).

The Case

DESCRIPTION

- (1) This patient is an 11 year old Caucasian boy. He weighs 62 pounds and is 53 inches tall. His presenting problem is that a physician doing a school physical three months ago, heard a murmur. T3: Now he is well developed, ok, that's normal. I guess most of these kids when they are referred to ah, pediatric cardiologist after a murmur is heard on school physical, most of them turn out to have a functional murmur. And, ah, and so my hunch right now is that he's got a functional

heart murmur. If I were to take a cross-section of these kids.

D1: There is nothing codable in this statement; that is no ValvAS, SubAS, SupAS.

HISTORY

- (2) The child was born following a normal pregnancy and delivery. He was delivered at full term with a birthweight of five pounds, six ounces.

T3: No comment.

D2: Nothing.

- (3) A murmur was heard at birth by the attending physician, but there has never been a follow-up of this murmur until the present.

T3: Again, that may be helpful but ah, again it could be a fluke. The kid can still have a functional murmur. It does lend credence to the fact that probably the kid may have a congenital heart lesion.

D3: Nothing.

- (4) Other than the murmur at birth, there were no problems in the neonatal period. Specifically, the child was not cyanotic in the neonatal period.

T3: No comment.

D4: Nothing.

- (5) The boy led a normal, asymptomatic childhood until the age of 10 years when, during an examination for a school physical, a murmur was noted by the physician doing the exam and he referred the child to the University of Minnesota for evaluation.

T3: I guess that was the last physical a year ago and it might take that long to get the kid plugged into the system.

D5: Nothing.

- (6) His growth and development have been normal when compared to his siblings and he has done well in school.

T3: No comment.

D6: Nothing.

- (7) There is no history of cyanosis or chest pains. He currently participates in baseball, football, and other sports and has no evidence of fatigue or dyspnea on exertion.

T3: Well, I guess we're being led down that whatever he's got is probably not significant. It is not a lesion that is causing him significant distress, either cyanosis or congestive heart failure or cardiac ah, limitation on a

cardiac basis. I guess would go along with mild aortic stenosis or subvalvular aortic stenosis, coarctation of the aorta, functional heart murmur, anything like that.

D7: "Aortic stenosis" (unqualified as to anatomical type) and "subvalvular aortic stenosis" are mentioned for the first time and both with a positive valence. Hence, a "T⁺" for both ValvAS and SubAS. This is a good example of why unqualified "aortic stenosis" is coded as valvular. The subject contrasts "aortic stenosis" with SubAS. When a subject means to refer to a non-valvular variant, he will do so. This may not be a perfect heuristic but it is a good one.

- (8) There have been no unusual childhood illnesses, and he has had no hospitalization or operations.

T3: You know this kid still could have acquired heart disease and ah, I guess that he may have had subclinical SBE, this happens a lot, that kids have got a subclinical case of SBE and are cured and show up and have a murmur of ASAI and then nothing else. The kid could have had a bicuspid valve at birth and a subclinical case of SBE I suppose. Whatever it is, it isn't

causing him much trouble.

D8: Even though it is conjoined with AI, that is, ASAI, Aortic Stenosis is mentioned here and not further qualified: hence, score ValvAS. "Bicuspid Aortic Valve" is mentioned here and we are scoring this as ValvAS. Both mentions of ValvAS are mildly positive; hence, score "+" for ValvAS.

- (9) The family history is negative for congenital heart disease.

T3: No comment.

D9: Nothing.

- (HHX) Hunches after history, please.

T3: My hunch at this point is that the kid has got minimal congenital heart disease, right now. And he probably has low-grade aortic stenosis, maybe he has subvalvular aortic stenosis which is a common lesion, he may have coarctation of the aorta. Those are my hunches following history.

DHHX: Score "+" for both ValvAS and SubAS.

PHYSICAL EXAM

- (10) On physical examination, this is a healthy appearing, 11 year old boy. The facies are normal.

T3: Ah, no comment.

D10: Nothing.

- (11) There is no cyanosis and the fingers and toes appear normal.

T3: No comment.

D11: Nothing.

- (12) There is no evidence of respiratory distress. Respiration rate is 16/minute.

T3: No comment.

D12: Nothing.

- (13) The blood pressure in the left arm is 90/70, in the right arm 102/60 and in the leg is 120/60 by auscultation.

T3: Ah, I think in kids that you can have ah, ah, blood pressure in your lower extremities higher than your right, than your upper extremities. I think that the maximum you are allowed though is ah, the maximum is 15 millimeters of mercury which this kid exceeds. I wonder if he has got multiple peripheral artery stenoses which we've seen a couple of cases of. And that would go along again with bicuspid aortic valve just like coarctation of the aorta. I think probably that this might be a normal finding, let's see what his pulses are.

D13: "Bicuspid aortic valve" is the only thing

here for which we are scoring; hence, "+" for ValvAS.

- (14) The pulses are normal and equal in arms and legs. The pulse rate is 84/minute.

T3: Ah, I'm inclined to believe on the basis of this, that the kid has got higher blood pressure in his legs is a normal finding. Especially if they did it with the wrong size cuffs, we're not told about that. It might be a normal finding.

D14: Nothing.

- (15) The chest is symmetrical with no bulge. There is mild asymmetry of the back associated with slight thoracic scoliosis. The lungs are clear to auscultation.

T3: Now I assume that this kid didn't have scoliosis when he was a baby. But, kids with scoliosis have a variety of heart murmurs which may or may not be associated with congenital heart disease. On the other hand kids with congenital heart disease frequently have some scoliosis. That's all.

D15: Nothing.

- (16) The apex impulse is in the fifth intercostal space in the mid-clavicular line.

T3: That's normal.

D16: Nothing.

- (17) There is a systolic thrill felt below the right clavicle, along the mid-left sternal border, and in the suprasternal notch.

T3: That would go along with aortic stenosis, I think. They don't say this kid has got any other kind of vascular disease like Marfans or anything like that that might present with multiple artery stenoses, with multiple peripheral artery stenoses or diffuse arterial disease. Let me reserve the right until I see his x-rays and junk in here, and see the rest of his exam. But I think this kid has probably got aortic stenosis on the basis of the suprasternal notch thrill. Ah, okay, ah, I think I'm at 18, right?

E: Yeah, you're on 18.

D17: Unqualified "aortic stenosis" is mentioned twice, both positively. Hence code "+" for ValvAS.

- (18) The first heart sound is normal and there is no systolic ejection click.

T3: I guess I'm led to then, to the notion that this kid has not got ah, valvular aortic stenosis. Frequently kids with supra-aortic stenosis may have a thrill, so mentioned,

but also these kids may have unusual facies and they have cupid's bow mouth and saddle nose, and, no, unusual nose, kind of cutsey-pie elfin facies. Ah, a lot like the kids with supravulvular pulmonary artery stenosis. But you don't need to have that. Subvalvular aortic stenosis may give the kid a suprasternal notch, ah, thrill, but ah, I'm not ah, persuaded by what I've seen so far. Okay, still this kid could have ah, and still this kid could have ah, arterial disease, pre-existing arterial disease.

D18: ValvAS, SubAS, and SupAS are all mentioned here. There is only one mention of Valvular AS and it is surely negative. "Supravulvular aortic stenosis" is mentioned for the first time and should be coded with a "T" at 18 even though the functional data item for the triggering seems to be 17. Deciding for the valence for the triggering of SupAS here is difficult. I judged this to be a case of "multiple evaluations that cancel each other out":

"Kids with SupAS may have a thrill so mentioned"

"but also these kids may have unusual facies"

"but you don't need to have that"

Not every such constellation of statements need be coded "o". Every effort should be made to decide an overall, net valence (+, -). I simply could not decide either way on SupAS in this instance and coded it "T^O". "Subvalvular aortic stenosis" is mentioned and I judged its valence to be mildly positive.

Note, the alternative scorer judged the evaluative valence for SubAS here to be neutral (o). She took the reference to be one of "multiple evaluations that cancel":

"SubAS may give a thrill"

"but I'm not persuaded by what I've seen so far"

I took the same set of statements to mean "SubAS is possible but what I've seen so far is not enough to conclude this." Hence, I saw the statement as mildly positive at this time.

- (19) A grade 4 over 6 systolic ejection murmur is heard best at the right upper sternal border, but also heard along the left sternal border, and less well throughout the precordium. The murmur is also heard in the neck.

T3: Aortic stenosis.

D19: "Aortic stenosis", hence, "+" on ValvAS.

This particular coding is worth note. It may well be the case that this particular subject, and others like him who obviously have knowledge of other AS variants and have used them repeatedly earlier in the case, may in fact mean more by a general "aortic stenosis" statement than just Valvular AS; it may well be a "shorthand" for all the aortic stenoses. However, across all of these subjects and differing particular mentions of unqualified "aortic stenosis", trying to untangle different meanings for the same words, i.e., "aortic stenosis", would be too unreliable and tenuous a process. Hence, we will stay "close to the belt". Whenever "aortic stenosis" and its correspondents (first sheet of instructions for scoring Table 1, Appendix F) is given without qualification as to more specific location (supra or sub), the only safe, consistent thing to do is code it as ValvAS. This will be right most of the time.

- (20) The second heart sound is normal and followed by a grade 1-2 over 6 early diastolic murmur heard best at the second left intercostal space.

T3: Aortic insufficiency. So this kid's got

ASAI, no click, he's probably got subvalvular aortic stenosis.

D20: This one is difficult. "Subvalvular aortic stenosis" is mentioned and the tone is positive: hence, "+" for SubAS. The hard part is what to do with the unqualified "AS" as part of "ASAI". By the rules, this should be treated as ValvAS. In keeping with the rules, I coded a mention of ValvAS and coded the valence as "o".

(21) Neither the liver nor the spleen was palpable.

T3: Ah, I think I don't have any comment about that. And they don't say anything about family history.

D21: Nothing.

(HPEX) At the end, do you have hunches at the end of this?

T3: Yeah, I do. I think this kid has got subvalvular aortic stenosis probably membranous. The only thing that they don't mention anything about what the murmur does with exercise or changing in position. But I think that this kid has got a history that is most like ah, membranous subaortic stenosis.

DHPEX: "Subvalvular aortic stenosis",
"probably membranous", and "membranous sub-
aortic stenosis" are all input to one "+" on
SubAS.

X-RAY

(22) The chest X-ray shows normal cardiac size and
contour and normal vascularity, but prominence
of the ascending aorta.

T3: The ah, this would go along with ah, with
ah, membranous subaortic stenosis, as well.

I'm, okay, it would go along with membranous
subaortic stenosis as well. If he has some
dilatation of the aortic root it could be on
the basis of his ASAI and that ASAI is probably
secondary to the wear and tear on the valve.
Usually, __ says that most kids with Valvular AS
have, valvular AS have a click and its absence
mitigates against valvular aortic stenosis.

The kid has probably got a subvalvular mem-
brane and AI secondary to that, secondary to
that. And that is probably accounting for his
secondary findings.

D22: The statements regarding SubAS are clearly positive and the statements regarding "Valvular aortic stenosis" are clearly negative: "absence (of click) mitigates against valvular aortic stenosis".

Note, this is a nice example of how a subject reconciles AI in SubAS, that is, as resulting from secondary damage to the aortic valve from blood "pounding" on it over the sub-aortic stenosis (see Scoring Guidelines for Table 1, Appendix F).

(HXray) Hunches after X-ray?

T3: My hunch at the end of x-ray is the same, that he has, that he has ah, membranous Sub-aortic Stenosis with aortic valvular insufficiency.

DHXray: Positive SubAS is clear.

EKG:

(23) The EKG shows a QRS axis of between $+30^{\circ}$ and $+45^{\circ}$. The P waves, T waves, and PR interval are normal. The pattern shows probable left ventricular hypertrophy.

T3: So he's not got strain. He's not got, if he doesn't have an ectopic T wave axis he's not got strain, his axis is normal but

in the leftward direction he's got probably left ventricular hypertrophy, he's probably got some mild aortic outflow obstruction, he probably does have a subvalvular membrane, he probably does have aortic insufficiency ah, which is probably moderate, probably not inconsiderable and he probably doesn't have aortic valvular stenosis. If he did have a click I would say that this kid probably had a bicuspid aortic valve, occult secondary ah, subacute bacterial endocarditis, now resolved. I suppose that the absence of a click may be, ah, may be misleading so that I guess that is probably my secondary diagnosis. Is that okay? My summary then, do you want?

D23: "He probably does have a subvalvular membrane . . . and he probably doesn't have aortic valvular stenosis." A plus for SubAS and a minus for ValvAS seems clear. Note that the deciding issue is really the "no click".

(HEKG) Yeah, why don't you just give me a set of hunches now and then I'll ask you ah, you know, the straightforward question.

T3: Okay, my hunch is that the kid has got subvalvular AS with ah . . . subvalvular

membranous AS with secondary aortic insufficiency, that's my hunch.

DHEKG: Positive SubAS

E: Okay, now we can get to the other thing. I would like you to, as explicitly as possible, give me a primary diagnosis.

T3: My primary diagnosis is subvalvular membranous aortic stenosis with aortic insufficiency.

E: Okay.

T3: And my secondary diagnosis is ah, is ah, aortic stenosis with insufficiency ah, probably with bicuspid aortic valve and probably ah, with a history of occult subacute bacterial endocarditis.

Comment

Note here that in an obvious reference to Valvular Aortic Stenosis, the subject uses the simple term "aortic stenosis". ValvAS is surely the "unmarked case" or "default value" of unqualified "aortic stenosis".

APPENDIX I

SCORING INSTRUCTIONS

Subvalvular Aortic Stenosis Instructions for Scoring Target Data Items

The object of this scoring is to document how the subject evaluates each of a set of target data items with respect to the major variants of Aortic Stenosis, that is, Valvular Aortic Stenosis (or unqualified "Aortic Stenosis"), Subvalvular Aortic Stenosis, and Supravalvular Aortic Stenosis. Generally speaking, the target data items are the main object of this scoring and the intent is to determine whether the subject interprets each target item as positive, negative, or neutral evidence for the presence in the patient of each Aortic Stenosis variant.

There are six target data items. Some items constitute the entire numbered data item presented to the subject in the case; others are a subpart of a numbered data item. The six target items are given below. For those that are a subpart of a numbered data item, the target subpart is underlined.

NORMAL FACIES:

- (10) On physical examination, this is a healthy-
appearing, 11 year old boy. The facies

are normal.

THE THRILL:

- (17) There is a systolic thrill felt below the right clavicle, along the mid-left sternal border, and in the suprasternal notch.(ENTIRE)

NO CLICK:

- (18) The first heart sound is normal and there is no systolic ejection click.

THE MURMUR:

- (19) A grade 4 over 6 systolic ejection murmur is heard best at the right upper sternal border, but also heard along the left sternal border, and less well throughout the precordium. The murmur is also heard in the neck.(ENTIRE)

AORTIC INSUFFICIENCY:

- (20) The second heart sound is normal and followed by a grade 1-2 over 6 early diastolic murmur heard best at the second left intercostal space.

PROMINENT AORTA:

- (22) The chest x-ray shows normal cardiac size and contour and normal vascularity, but prominence of the ascending aorta.

As with Table 1 scoring, the target set of diseases is the set of Aortic Stenosis variants: Valvular AS, Subvalvular AS, and Supravalvular AS. The set of

correspondences between these three types of AS and particular subtypes and wordings that are to be included within each is the same as for Table 1 scoring and the first page of Table 1 scoring instructions should be consulted.

Evaluation of a target item with respect to one or more of the AS variants need not occur at the time when a target data item is presented (e.g., a subject may evaluate 17 at time of presentation of, say 18). The objective is to score the evaluation wherever it occurs. Preliminary screening has shown that the only evaluations of the target items in relation to AS variants occur at item 10 and items 17-HEKG. Hence, scoring will focus on subjects' responses at these spots which are potential "places of evaluation" (POE's).

Scoring in Table 2 really involves three kinds of decisions:

- (a) Deciding if any and which Aortic Stenosis variant(s) is mentioned.
- (b) For AS variants mentioned, deciding if any and which target item(s) is being evaluated.
- (c) Deciding the valence (+, -, o) for the evaluation.

Scoring Procedure

A table (Table 2) is provided for this scoring. Across the top are the six target items and, under each, codes for each of the three AS variants. On the left margin, for each subject, are listed the ten potential places of evaluation (POE'S) which are ordered from top to bottom (10, 17-HEKG).

Read the entire subject response at each POE, taken in order, starting at 10 and then skipping to 17. At each POE, decide if any and which AS variants are mentioned. For each AS variant mentioned, decide if any and which target data items are being evaluated with respect to the AS variant. For each AS variant - target data item pair, decide the valence of the evaluation (positive, negative, neutral) and place a "+", "-", or "o" to the right of the POE and in the column corresponding to the AS variant and target data item being evaluated by the subject.

Guidelines and Cautions

I. Deciding if any and which AS variant is mentioned

This component of the task of scoring Table 2 is the same as the scoring task for Table 1 and the correspondence conventions (first page, Table 1

scoring, Appendix F) and scoring guidelines (especially Table 1, #IV, Appendix F, Levels of ambiguity in referencing AS variants) for that table should be applied. In particular, the AS variant being evaluated by the subject in relation to a target data item will not always be named explicitly. Table 2 scoring is quite important; hence, every attempt should be made to disambiguate which AS variant is being evaluated. The Table 1 guidelines set reasonable boundaries on how far we can go in attempting disambiguation.

For Table 2, it is not enough that an AS variant be mentioned; that is only the first step in deciding a codable item. In addition, the AS variant, once mentioned, must be evaluated with respect to some target data item.

II. Deciding if any and which target item(s) is evaluated

To constitute something scorable, an AS variant must be evaluated with respect to some target data item. Evaluation of an AS variant with respect to a target item means that a target item is referred to by the subject, in relation to an AS variant, in a manner that indicates that the target item bears on the status of the AS variant as a candidate hypothesis

(i.e., the target item is evaluated as evidence, of whatever kind, for the AS variant).

A difficulty arises because some target data items are subparts of numbered data items (presented in the case) and we are only interested in the target subpart (as discussed earlier in the Appendix). The following rules should apply.

Target subpart mentioned explicitly: If the target subpart of numbered data item is explicitly mentioned (e.g., "no click") in an evaluation, there is no problem. Score the evaluation.

Numbered data item only: If only the numbered data item containing the target subpart is referred to in an evaluation - without distinction as to which subpart is being evaluated - score it as though it were the target subpart. In this sense, the target subpart is the "default" for the numbered data item. That is, unless otherwise indicated, we assume the subject is referring to the target.

Non-target subpart mentioned explicitly: When the non-target subpart of a numbered data item is referred to explicitly in an evaluation, the evaluation is not to be scored.

Eg. (18) The first heart sound is normal and there is no systolic ejection click.

S: (The first heart sound is normal so probably the two A-V valves are normal). There is no systolic ejection click as you would expect with valvular aortic stenosis . . .

The first parenthesized statement, since it explicitly refers to the non-target part of 18, is not a scorable statement (even if the subject had evaluated the "normal first heart sound" with regard to some AS variant). The remainder of this subject's statement is scorable.

III. Judging +, -, o for the evaluation

Whenever an AS variant is evaluated with respect to a target data item, the valence of the evaluation can run from extreme negative ("Well, that rules out SubAS.") to extreme positive ("Well, that strongly supports the idea that we are dealing with SubAS in this patient") with some turning point in between. The general guidelines given in the instructions for Table 1, #II, Appendix F, should be followed here.

Again, it should be emphasized that the zero convention should be used sparingly; use the subject's whole statement and make every effort to detect a positive or negative valence. The zero convention should be reserved for the instances where you simply cannot

decide.

"Multiple evaluations that cancel each other out" is still a good candidate for the zero convention. (But not necessarily - only if you cannot decide the net valence). However, in Table 2 scoring, this would refer only to multiple evaluations at the same POE. If a subject makes an additional evaluation of a target item at a different POE, score the POE's separately. This is the main reason for including different POE's in the table.

IV. Scoring at "hunches" points

Remember, that the objective of Table 2 scoring is to score evaluations of target data items. Hence, when mention is made of AS variants at "hunches" points, they are only scorable if they refer to some target data item. Subject discussions of AS variants at "hunches" points that do not make use of some target data item are not of interest here.

V. Target items that happen to be trigger points

It makes no difference, for this table, if a target item happens to be the item where an AS variant is first mentioned. In the sense of this table, triggering is an evaluation like any other. Just decide the positive, negative, or neutral valence of the target

item - AS variant relationship.

VI. Target item 20 - Aortic Insufficiency

Data item 20 is a straightforward description of the murmur of Aortic Insufficiency (AI) and just about all subjects will interpret it as such. This itself is not the interpretation of interest in this table. The question, for this table, is whether the subject then goes on to interpret either AI or 20 in relation to, that is, as evidence for, a variant of AS. The following guidelines should apply.

- (a) If the subject interprets 20 as AI but makes no evaluation of AI with respect to an AS variant, the statement is not scorable.
- (b) If the subject interprets 20 as AI and evaluates AI with respect to some AS variant, the statement is scorable.
- (c) If the subject does not explicitly interpret 20 as AI, but does evaluate 20 with respect to some AS variant, score the statement. In this instance, we will assume the evaluation of 20 is mediated by AI.

AI is a condition that occurs more often with some variants of AS than with others; hence, AI constitutes legitimate evidence for deciding which variant of AS is

present in the patient. We are only interested in AI when it is used by the subject for this function.

VII. AI and the "prominent aorta"

In evaluating "prominent aorta" with regard to Subvalvular AS, a subject may make a statement that, on the surface of it, seems negative. If, however, he goes on to say that the "prominent aorta" could be accounted for, under SubAS, when SubAS is accompanied by AI, then you should try to decide if the subject is using the accompanying AI to account for the aorta. If you decide he is, then the real evaluation is more likely positive. This is a judgment call and is left to the scorer. The subject's response to 20 can provide some guidance by providing evidence of how convinced the subject is that AI is present. The following is an example that appears to take AI into account ("valvar damage" means AI):

Eg. (22) The chest X-ray shows normal cardiac size and normal vascularity, but prominence of the ascending aorta.

S: Again, this is consistent with aortic stenosis, ah, and this could represent just post-stenotic dilatation. Ah, again, usually you don't have, you don't have, post-stenotic dilatation with SubAs unless you

have some valvar damage as well. That is about all I can say.

This is surely positive for general Aortic Stenosis. The questionable part is the evaluation with respect to SubAS. The evaluation should probably be taken as having a positive valence, especially since the subject had earlier (at 20) strongly tied SubAS to AI:

(20) _____

S: Now, this is consistent with some aortic insufficiency and, ah, adds weight to the possibility of a subaortic membrane ah, which has a high incidence of a diastolic murmur associated with it.

VIII. Hypothetical considerations of data

At times, subjects make statements about what their thoughts might be, had the data been other than it was. We can only handle data that was presented in the case and cannot deal with what might have been:

Eg. (HEKG)

S: So my continuing hunch is congenital aortic stenosis, presumably subaortic if the examiner is correct in his description of the physical

findings, but if there is an ejection click present, that was missed by the examiner, then this is congenital valvular aortic stenosis with some hypertrophy.

The case, in fact, presented "no click" which is judged here as positive for SubAS. The subject's proposal of the presence of a click and its positive loading on Valvular AS is hypothetical.

APPENDIX J

CASE 1

Coding by the Second Scorer for Table 2

The purpose of coding that produced Table 2 in the report body for Case 1, was to document the evaluation by subjects of six target data items with respect to the variants of Aortic Stenosis. Subjects were free to evaluate, at any later time, any data item they had already seen. Hence, evaluations of a target item were not restricted to the place in the case where the item was first presented to the subject. The objective in coding for Table 2 was to capture all evaluations of a target item as positive(+), negative (-), or "neutral" (o) evidence for the existence of one of the variants of Aortic Stenosis in the patient, no matter when these evaluations took place ("Zero" is best thought of as an evaluation for which the scorer could not decide whether the data item was viewed as positive or negative evidence by the subject - see Appendix I for scoring instructions on this point).

When a scorer judged that there was more than one separate evaluation of a target item with respect to the same AS variant, the "net" evaluation was used in constructing Table 2. Neither scorer judged that there

were ever more than two separate evaluations of any target data item with respect to the same variant; hence, the following rule was used in constructing Table 2:

<u>First Evaluation</u>	<u>Second Evaluation</u>	<u>Recorded in Table 2</u>
+	+	+
+	-	0
-	-	-

No combination involving two uses of "zero" occurred for either scorer.

Table 1-J gives the equivalent of Table 2 produced from the coding of the second scorer. A non-blank cell entry in either of these two tables indicates that the scorer judged a data item - AS variant evaluation (at least one) to have taken place. There are a combined total of 88 such entries across the two tables. The two scorers agreed on 83 of these (94%) that a codable evaluation had taken place. The cells of disagreement are marked with a single "check" in Table 1-J indicating that the author judged an evaluation to have occurred while the second scorer did not or vice-versa. There are 83 data item - AS variant cells in which both scorers judged an evaluation to have taken place. In 78 of these (94%), the two scorers assigned the

Table 1-J
Case 1: Subvalvular Aortic Stenosis: Evaluations of Target Data
Items with Respect to Members of the LCS

	(10) Normal Facies		(17) Thrill		(18) No Click		(19) Murmur		(20) Aortic Insuff		(22) Prominent Aorta	
	ValvAS	SubAS	SupAS	ValvAS	SubAS	SupAS	ValvAS	SubAS	SupAS	ValvAS	SubAS	SupAS
Students												
S1	-				-		+				+	
S2	+J	0J	-	+	J		+		+		+	
S3				+	+		+		+		+	
S4							+		+J		+	
Trainees												
T1				+	-		+	+	+		+	0/J
T2							+				+	
T3	0			+	-	+	+		+			+
T4				+	0/J		+		+J		+	
Experts												
E1	-		+	0	+	+	+	0	+	+	+	-
E2	-				0/J	+	+		+			
E3*	-		+	+	0/J	+	+	+	+	+	+	0/J
E4*			+		-	+	+	+	+	+	+	+

Note. (J) indicates cell where scorers disagreed that an evaluation had taken place.
(J J) indicates cell where scorers disagreed on the value of an evaluation.
* The subjects with more than twenty years of experience.

identical value (+, -, o) to the evaluation. No disagreements crossed the "zero boundary"; that is, there is no cell where one scorer assigned a "plus" while the other assigned a "minus". The cells in which the two scorers disagreed on the "value" of the evaluation are marked with a double "check" in Table 1-J.

Except for the disagreements on the evaluations of "NO CLICK" in relation to Valvular Aortic Stenosis, no other disagreements would in any way alter the analyses or interpretation based on Table 2 in the body of the report except that one student (S2), based on Table 1-J, would be given "credit" for evaluating all three variants with respect to one (the same) data item.

The only effect of the "NO CLICK" disagreements would be to make the discussion of the evaluations of this finding by some individual subjects somewhat less "clean" (although it would not be altered in any substantial way). This being the case, the disagreements on this finding merit some discussion.

The author judged S2 to have evaluated "NO CLICK" as positive for ValvAS while the second scorer judged no evaluation to have taken place. The problem here is an indirect referent: "The fact that there is no systolic ejection, that would go along too . . . that

would again go along with what I'm thinking." The question is whether a referent can be assigned for what the subject is "thinking." At the immediately preceding data item the subject had said that his "prime possibility" was Valvular Aortic Stenosis. Scorers were allowed to assign indirect referents if they were confident in them (See Appendix F instructions) and this referent seemed clear to the author.

Regarding T4, the relevant protocol for the subject is given as an example in Figure 5 of the body of the report. This is a relatively clear negative evaluation by the standards of what one sees in the protocols from this study, although the author may have benefitted from having heard the consternation in the tone of the subject's voice at this point in the case.

The difficulty for both subjects E2 and E3 has exactly the same basis which can be seen in E3's protocol given in Figure 4 of the body of the report. The subjects raised the idea that Valvular Aortic Stenosis of a "very severe degree" might present without a click. The author took this as a "hypothetical" (i.e., "no click" would be expected had the AS been severe) and judged the basic evaluation to be negative.

Navy

- 1 Dr. Ed Aiken
Navy Personnel R&D Center
San Diego, CA 92152
- 1 Meryl S. Baker
NPRDC
Code P309
San Diego, CA 92152
- 1 Dr. Robert Breaux
Code N-711
NAVTRAEQUIPCEN
Orlando, FL 32813
- 1 CDR Mike Curran
Office of Naval Research
800 N. Quincy St.
Code 270
Arlington, VA 22217
- 1 DR. PAT FEDERICO
NAVY PERSONNEL R&D CENTER
SAN DIEGO, CA 92152
- 1 Dr. John Ford
Navy Personnel R&D Center
San Diego, CA 92152
- 1 LT Steven D. Harris, MSC, USN
Code 6021
Naval Air Development Center
Warminster, Pennsylvania 18974
- 1 Dr. Jim Hollan
Code 304
Navy Personnel R & D Center
San Diego, CA 92152
- 1 Dr. Norman J. Kerr
Chief of Naval Technical Training
Naval Air Station Memphis (75)
Millington, TN 38054
- 1 Mr. Orvin A. Laron
Code 306
Navy Personnel R&D Center
San Diego, CA 92152

Navy

- 1 Dr. William L. Maloy
Principal Civilian Advisor for
Education and Training
Naval Training Command, Code 00A
Pensacola, FL 32508
- 1 CAPT Richard L. Martin, USN
Prospective Commanding Officer
USS Carl Vinson (CVN-70)
Newport News Shipbuilding and Drydock Co
Newport News, VA 23607
- 1 Dr William Montague
Navy Personnel R&D Center
San Diego, CA 92152
- 1 Ted M. I. Yellen
Technical Information Office, Code 201
NAVY PERSONNEL R&D CENTER
SAN DIEGO, CA 92152
- 1 Library, Code P201L
Navy Personnel R&D Center
San Diego, CA 92152
- 1 Technical Director
Navy Personnel R&D Center
San Diego, CA 92152
- 6 Commanding Officer
Naval Research Laboratory
Code 2627
Washington, DC 20390
- 1 Psychologist
ONR Branch Office
Bldg 114, Section D
666 Summer Street
Boston, MA 02210
- 1 Psychologist
ONR Branch Office
536 S. Clark Street
Chicago, IL 60605
- 1 Office of Naval Research
Code 437
800 N. Quincy Street
Arlington, VA 22217

Navy

- 5 Personnel & Training Research Programs (Code 458)
Office of Naval Research
Arlington, VA 22217
- 1 Psychologist
ONR Branch Office
1030 East Green Street
Pasadena, CA 91101
- 1 Special Asst. for Education and Training (OP-01E)
Rm. 2705 Arlington Annex
Washington, DC 20370
- 1 Office of the Chief of Naval Operations Research Development & Studies Branch (OP-115)
Washington, DC 20350
- 1 LT Frank C. Petho, MSC, USN (Ph.D)
Selection and Training Research Division
Human Performance Sciences Dept.
Naval Aerospace Medical Research Laborat
Pensacola, FL 32508
- 1 Dr. Gary Poock
Operations Research Department
Code 55PK
Naval Postgraduate School
Monterey, CA 93940
- 1 Roger W. Remington, Ph.D
Code L52
NAMRL
Pensacola, FL 32508
- 1 Dr. Worth Scanland, Director
Research, Development, Test & Evaluation
N-5
Naval Education and Training Command
NAS, Pensacola, FL 32508
- 1 Dr. Robert G. Smith
Office of Chief of Naval Operations
OP-987H
Washington, DC 20350

Navy

- 1 Dr. Alfred F. Snode
Training Analysis & Evaluation Group
(TAEG)
Dept. of the Navy
Orlando, FL 32813
- 1 Dr. Richard Sorensen
Navy Personnel R&D Center
San Diego, CA 92152
- 1 Roger Weissinger-Baylon
Department of Administrative Sciences
Naval Postgraduate School
Monterey, CA 93940
- 1 Dr. Robert Wisher
Code 309
Navy Personnel R&D Center
San Diego, CA 92152
- 1 Mr John H. Wolfe
Code P310
U. S. Navy Personnel Research and
Development Center
San Diego, CA 92152

Army

- 1 Technical Director
U. S. Army Research Institute for the
Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333
- 1 Mr. James Baker
Systems Manning Technical Area
Army Research Institute
5001 Eisenhower Ave.
Alexandria, VA 22333
- 1 Dr. Beatrice J. Farr
U. S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333
- 1 Dr. Dexter Fletcher
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333
- 1 DR. FRANK J. HARRIS
U.S. ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333
- 1 Dr. Michael Kaplan
U.S. ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333
- 1 Dr. Milton S. Katz
Training Technical Area
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333
- 1 Dr. Harold F. O'Neil, Jr.
Attn: PERI-OK
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333
- 1 Dr. Robert Sasmor
U. S. Army Research Institute for the
Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

Army

- 1 Dr. Frederick Steinheiser
Dept. of Navy
Chief of Naval Operations
OP-113
Washington, DC 20350
- 1 Dr. Joseph Ward
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Air Force

- 1 Dr. Earl A. Alluisi
HQ, AFHRL (AFSC)
Brooks AFB, TX 78235
- 1 Dr. Genevieve Haddad
Program Manager
Life Sciences Directorate
AFOSR
Bolling AFB, DC 20332
- 2 3700 TCHTW/TTGH Stop 32
Sheppard AFB, TX 76311

Marines

- 1 H. William Greenup
Education Advisor (E031)
Education Center, MCDEC
Quantico, VA 22134
- 1 Special Assistant for Marine
Corps Matters
Code 100M
Office of Naval Research
800 N. Quincy St.
Arlington, VA 22217
- 1 DR. A.L. SLAFKOSKY
SCIENTIFIC ADVISOR (CODE RD-1)
HQ, U.S. MARINE CORPS
WASHINGTON, DC 20380

CoastGuard

- 1 Chief, Psychological Research Branch
U. S. Coast Guard (G-P-1/2/TP42)
Washington, DC 20593

Other DoD

- 12 Defense Technical Information Center
Cameron Station, Bldg 5
Alexandria, VA 22314
Attn: TC
- 1 Military Assistant for Training and
Personnel Technology
Office of the Under Secretary of Defense
for Research & Engineering
Room 3D129, The Pentagon
Washington, DC 20301
- 1 DARPA
1400 Wilson Blvd.
Arlington, VA 22209

Civil Govt

- 1 Dr. Paul G. Chapin
Linguistics Program
National Science Foundation
Washington, DC 20550
- 1 Dr. Susan Chipman
Learning and Development
National Institute of Education
1200 19th Street NW
Washington, DC 20208
- 1 William J. McLaurin
66610 Howie Court
Camp Springs, MD 20031
- 1 Dr. Arthur Melmed
National Institute of Education
1200 19th Street NW
Washington, DC 20208
- 1 Dr. Andrew R. Molnar
Science Education Dev.
and Research
National Science Foundation
Washington, DC 20550
- 1 Dr. Joseph Psotka
National Institute of Education
1200 19th St. NW
Washington, DC 20208
- 1 Dr. Frank Withrow
U. S. Office of Education
400 Maryland Ave. SW
Washington, DC 20202
- 1 Dr. Joseph L. Young, Director
Memory & Cognitive Processes
National Science Foundation
Washington, DC 20550

Non Govt

- 1 Dr. John R. Anderson
Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213
- 1 Anderson, Thomas H., Ph.D.
Center for the Study of Reading
174 Children's Research Center
51 Gerty Drive
Champaign, IL 61820
- 1 Dr. John Annett
Department of Psychology
University of Warwick
Coventry CV4 7AL
ENGLAND
- 1 DR. MICHAEL ATWOOD
SCIENCE APPLICATIONS INSTITUTE
40 DENVER TECH. CENTER WEST
7935 E. PRENTICE AVENUE
ENGLEWOOD, CO 80110
- 1 1 psychological research unit
Dept. of Defense (Army Office)
Campbell Park Offices
Canberra ACT 2600, Australia
- 1 Dr. Alan Baddeley
Medical Research Council
Applied Psychology Unit
15 Chaucer Road
Cambridge CB2 2EF
ENGLAND
- 1 Dr. Patricia Baggett
Department of Psychology
University of Colorado
Boulder, CO 80309
- 1 Mr Avron Barr
Department of Computer Science
Stanford University
Stanford, CA 94305
- 1 Liaison Scientists
Office of Naval Research,
Branch Office, London
Box 39 FPO New York 09510

Non Govt

- 1 Dr. Lyle Bourne
Department of Psychology
University of Colorado
Boulder, CO 80309
- 1 Dr. John S. Brown
XEROX Palo Alto Research Center
3333 Coyote Road
Palo Alto, CA 94304
- 1 Dr. Bruce Buchanan
Department of Computer Science
Stanford University
Stanford, CA 94305
- 1 DR. C. VICTOR BUNDERSON
WICAT INC.
UNIVERSITY PLAZA, SUITE 10
1160 SO. STATE ST.
OREM, UT 84057
- 1 Dr. Pat Carpenter
Department of Psychology
Carnegie-Mellon University
Pittsburgh, PA 15213
- 1 Dr. John B. Carroll
Psychometric Lab
Univ. of No. Carolina
Davie Hall 013A
Chapel Hill, NC 27514
- 1 Dr. William Chase
Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213
- 1 Dr. Micheline Chi
Learning R & D Center
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15213
- 1 Dr. William Clancey
Department of Computer Science
Stanford University
Stanford, CA 94305

Non Govt

- 1 Dr. Allan M. Collins
Bolt Beranek & Newman, Inc.
50 Moulton Street
Cambridge, Ma 02138
- 1 Dr. Lynn A. Cooper
LRDC
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15213
- 1 Dr. Meredith P. Crawford
American Psychological Association
1200 17th Street, N.W.
Washington, DC 20036
- 1 Dr. Kenneth B. Cross
Anacapa Sciences, Inc.
P.O. Drawer Q
Santa Barbara, CA 93102
- 1 LCOL J. C. Eggenberger
DIRECTORATE OF PERSONNEL APPLIED RESEARC
NATIONAL DEFENCE HQ
101 COLONEL BY DRIVE
OTTAWA, CANADA K1A 0K2
- 1 Dr. Ed Feigenbaum
Department of Computer Science
Stanford University
Stanford, CA 94305
- 1 Dr. Richard L. Ferguson
The American College Testing Program
P.O. Box 168
Iowa City, IA 52240
- 1 Mr. Wallace Feurzeig
Bolt Beranek & Newman, Inc.
50 Moulton St.
Cambridge, MA 02138
- 1 Dr. Victor Fields
Dept. of Psychology
Montgomery College
Rockville, MD 20850

Non Govt

- 1 Dr. John R. Frederiksen
Bolt Beranek & Newman
50 Moulton Street
Cambridge, MA 02138
- 1 Dr. Alinda Friedman
Department of Psychology
University of Alberta
Edmonton, Alberta
CANADA T6G 2E9
- 1 DR. ROBERT GLASER
LRDC
UNIVERSITY OF PITTSBURGH
3939 O'HARA STREET
PITTSBURGH, PA 15213
- 1 Dr. Marvin D. Glock
217 Stone Hall
Cornell University
Ithaca, NY 14853
- 1 DR. JAMES G. GREENO
LRDC
UNIVERSITY OF PITTSBURGH
3939 O'HARA STREET
PITTSBURGH, PA 15213
- 1 Dr. Barbara Hayes-Roth
The Rand Corporation
1700 Main Street
Santa Monica, CA 90406
- 1 Dr. Frederick Hayes-Roth
The Rand Corporation
1700 Main Street
Santa Monica, CA 90406
- 1 Dr. Kristina Hooper
Clark Kerr Hall
University of California
Santa Cruz, CA 95060
- 1 Glenda Greenwald, Ed.
"Human Intelligence Newsletter"
P. O. Box 1163
Birmingham, MI 48012

Non Govt

- 1 Dr. Earl Hunt
Dept. of Psychology
University of Washington
Seattle, WA 98105
- 1 Dr. Ed Hutchins
Navy Personnel R&D Center
San Diego, CA 92152
- 1 Dr. Walter Kintsch
Department of Psychology
University of Colorado
Boulder, CO 80302
- 1 Dr. David Kieras
Department of Psychology
University of Arizona
Tucson, AZ 85721
- 1 Dr. Kenneth A. Klivington
Program Officer
Alfred P. Sloan Foundation
630 Fifth Avenue
New York, NY 10111
- 1 Dr. Stephen Kosslyn
Harvard University
Department of Psychology
33 Kirkland Street
Cambridge, MA 02138
- 1 Dr. Marcy Lansman
Department of Psychology, NI 25
University of Washington
Seattle, WA 98195
- 1 Dr. Jill Larkin
Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213
- 1 Dr. Alan Lesgold
Learning R&D Center
University of Pittsburgh
Pittsburgh, PA 15260

Non Govt

- 1 Dr. Erik McWilliams
Science Education Dev. and Research
National Science Foundation
Washington, DC 20550
- 1 Dr. Mark Miller
TI Computer Science Lab
C/O 2824 Winterplace Circle
Plano, TX 75075
- 1 Dr. Allen Munro
Behavioral Technology Laboratories
1845 Elena Ave., Fourth Floor
Redondo Beach, CA 90277
- 1 Dr. Donald A Norman
Dept. of Psychology C-009
Univ. of California, San Diego
La Jolla, CA 92093
- 1 Dr. Seymour A. Papert
Massachusetts Institute of Technology
Artificial Intelligence Lab
545 Technology Square
Cambridge, MA 02139
- 1 Dr. James W. Pellegrino
University of California,
Santa Barbara
Dept. of Psychology
Santa Barbara, CA 93106
- 1 MR. LUIGI PETRULLO
2431 N. EDGEWOOD STREET
ARLINGTON, VA 22207
- 1 DR. PETER POLSON
DEPT. OF PSYCHOLOGY
UNIVERSITY OF COLORADO
BOULDER, CO 80309
- 1 MINRAT M. L. RAUCH
P II 4
BUNDESMINISTERIUM DER VERTEIDIGUNG
POSTFACH 1328
D-53 BONN 1, GERMANY

Non Govt

- 1 Dr. Fred Reif
SESAME
c/o Physics Department
University of California
Berkeley, CA 94720
- 1 Mary Riley
LRDC
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15213
- 1 Dr. Andrew M. Rose
American Institutes for Research
1055 Thomas Jefferson St. NW
Washington, DC 20007
- 1 Dr. Ernst Z. Rothkopf
Bell Laboratories
600 Mountain Avenue
Murray Hill, NJ 07974
- 1 Dr. David Rumelhart
Center for Human Information Processing
Univ. of California, San Diego
La Jolla, CA 92093
- 1 Dr. Alan Schoenfeld
Department of Mathematics
Hamilton College
Clinton, NY 13323
- 1 DR. ROBERT J. SEIDEL
INSTRUCTIONAL TECHNOLOGY GROUP
HUMRRO
300 N. WASHINGTON ST.
ALEXANDRIA, VA 22314
- 1 Committee on Cognitive Research
% Dr. Lonnie R. Sherrod
Social Science Research Council
605 Third Avenue
New York, NY 10016

Non Govt

- 1 Robert S. Siegler
Associate Professor
Carnegie-Mellon University
Department of Psychology
Schenley Park
Pittsburgh, PA 15213
- 1 Dr. Edward E. Smith
Bolt Beranek & Newman, Inc.
50 Moulton Street
Cambridge, MA 02138
- 1 Dr. Robert Smith
Department of Computer Science
Rutgers University
New Brunswick, NJ 08903
- 1 Dr. Richard Snow
School of Education
Stanford University
Stanford, CA 94305
- 1 Dr. Robert Sternberg
Dept. of Psychology
Yale University
Box 11A, Yale Station
New Haven, CT 06520
- 1 DR. ALBERT STEVENS
BOLT BERANEK & NEWMAN, INC.
50 MOULTON STREET
CAMBRIDGE, MA 02138
- 1 David E. Stone, Ph.D.
Hazeltime Corporation
7680 Old Springhouse Road
McLean, VA 22102
- 1 DR. PATRICK SUPPES
INSTITUTE FOR MATHEMATICAL STUDIES IN
THE SOCIAL SCIENCES
STANFORD UNIVERSITY
STANFORD, CA 94305

Non Govt

- 1 Dr. Kikumi Tatsuoka
Computer Based Education Research
Laboratory
252 Engineering Research Laboratory
University of Illinois
Urbana, IL 61801
- 1 Dr. John Thomas
IBM Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598
- 1 Dr. Douglas Towne
Univ. of So. California
Behavioral Technology Labs
1845 S. Elena Ave.
Redondo Beach, CA 90277
- 1 Dr. Benton J. Underwood
Dept. of Psychology
Northwestern University
Evanston, IL 60201
- 1 Dr. Phyllis Weaver
Graduate School of Education
Harvard University
200 Larsen Hall, Appian Way
Cambridge, MA 02138
- 1 DR. GERSHON WELTMAN
PERCEPTRONICS INC.
6271 VARIEL AVE.
WOODLAND HILLS, CA 91367
- 1 Dr. Keith T. Wescourt
Information Sciences Dept.
The Rand Corporation
1700 Main St.
Santa Monica, CA 90406

END

DATE
FILMED

11-81

DTIC